The Journal *of*
Parapsychology

Volume 2 JUNE, 1938 Number 2

EDITORIAL COMMENT

REPORTS FROM Spring meetings of psychologists indicate that the critics of ESP research have had opportunity to air their views with their professional colleagues. In the New York meeting of the American Psychological Association three papers were presented. All were, to use *Time*'s expression, "hostile." The Southern Association for Philosophy and Psychology meeting at Knoxville devoted time to three papers, all reports of research claiming that ESP had been demonstrated under the respective conditions. The Mid-western meeting of the American Psychological Association included a symposium on ESP at which research reports, favorable and unfavorable, were discussed critically.

These are but some of the evidences that the ESP research has become a very live and active issue among psychologists, and that it continues to draw the fire of criticism. This is a healthy and welcome state of affairs for parapsychological research, which is already under especially great obligation to its critics and will long continue to need such assistance. No research is stronger than its weakest essential link, and its progress is greatly determined by the readiness with which the weaker points are detected and can be repaired.

In fairness to the investigators of extra-sensory perception, it must be recognized that the most pertinent and fruitful criticisms have come from the workers themselves. This might well be expected, since familiarity with the research situation is essential to relevant criticism. In this regard, the considerable expansion of the number of men and women experimenting in this field has been of inestimable value, bringing literally scores of new points of view to bear upon the problems.

But also from persons inactive in the research itself have there been important suggestions and criticisms made. These have come largely from those who have carefully followed the research reports, and these

criticisms have been for the most part made in a wholly constructive way. Particularly has this been true of the statistical aspect of the research field.

We may go one step further; the obligation extends also to those who, with some indication of unfamiliarity with the reports and with no evidence of any great concern to save the research from error, nevertheless launched attacks upon it which had the general effect of putting its investigators still more on their mettle. In spite of the fact that the great mass of this type of criticism proved to be irrelevant, it has served to stimulate considerable statistical research and to initiate some further experimentation along lines pointed out in criticism; it may be regarded as a wholesome influence even if only in an incidental way.

This is not, however, to say that all criticisms and all attacks have been fruitful, either to the investigation of ESP in particular or to science in general. Criticisms can be so irrelevant to the research and so misleading to readers not in close touch with the experimental work, as to create a wholly distorted view of its character and precautions. While any securely endowed and permanently established research can well withstand public misunderstanding and ridicule, this can hardly be regarded as true of any pioneer investigation, particularly if there is associated with the public misunderstanding the danger that young academic investigators may lose caste or professional opportunity by being associated with a research thus stigmatized.

There is, for example, no good purpose served by a gratuitous caricature of the research such as the photographically illustrated press article recently issued by a well-known science news agency. In this article the implication was that the explanation lay in transparent cards and shiny table tops that reflected the card symbols. The screened card tests and distance work upon which the case for ESP rests are so far from anything of this kind that a more misleading point to feature could hardly have been found. The fact too that a national news agency could give extremely wide publicity to this attack does not make the case any better. Likewise the showmanlike performance before the recent meeting of psychologists in New York in which ESP cards were called from sensory cues with the implication that this was the way the critical tests were given, can hardly be construed as of value, unless as entertainment.

However, this more questionable type of criticism has been, fortunately, a minor element; and perhaps its coming to the fore in recent

months may indicate that the argument against the case for extra-sensory perception is growing thin.

One significant feature stands out in the four years of criticism to which the ESP research has been subjected: Up to the present no critic has attempted a thorough and comprehensive evaluation of the research as a whole. Attention has been confined to possible weak points here and there which could not at best involve more than a minor portion of the total evidence or a slight alteration of the odds against chance. The rapid expansion of the explorations in this field make it particularly urgent that the primary and fundamental question of the occurrence of extra-sensory perception be established once and for all, or the essential weakness of the claim for it be pointed out by a very much more drastic analysis than has yet been forthcoming. It must be recognized at the same time that any criticism, in order to be completely destructive, must deal with all the various sources of evidence obtained under the most ideal experimental conditions, and with the now considerable number of independent repetitions. The situation offers a clear challenge to the serious critic to bring the entire research program to a halt as soon as possible if it can be done by a just and sound criticism. If this manifestly can not be done, we shall go on the more freely and rapidly to further advances for having this assurance of security behind us.

Assuming, however, that the case for an extra-sensory mode of perception will endure, we may safely anticipate for the future the need for a level of criticism well beyond the demands of the past research. For example, the problems raised when the attempt is made to arrange a crucial test of the hypothesis of precognition are extremely complex. The difficulties offered when clairvoyant and telepathic perception, unlimited by space, are regarded as counter-hypotheses are very great. A glimpse of this may be had in the article (No. II) on the precognition hypothesis in the present number, and this complication will increase as the series advances. There is some ground for fear that there may await us at some unforeseen point ahead, an experimental impasse resulting from the sheer inability to devise critical discriminatory tests. We shall have, then, only begun to need our critics.

* * *

One of the truisms in science is that no pioneer research is stronger than the confirmations it receives in other hands. While confirmatory repetitions of the ESP research have been far in the lead over failures to confirm, still the annoying fact has stood out that not everyone has

been able to find ESP ability in tests given presumably according to prescribed method.

Those engaged in the research who have had better success have not, of course, been free to suggest to those who fail that there were certain more subtle personal relationships that could be of importance in obtaining these results; but that this is so has been a general impression among experimenters. And now there comes to publication the report of a study which shows experimentally that such *personal factors are of the utmost importance.* The report of Pratt and Price will go far, first to suggest a most probable explanation for the failure of the minority of experimenters who have obtained negative results, and second, and much more important, to show a way out for these experimenters. It offers them a constructive suggestion—namely, that of getting help from someone else to create the favorable social environment for the subject during the test.

Considerable interest has been expressed from many quarters in the question as to whether any correlation is reliably indicated between ESP ability and other abilities or traits of personality. Such studies as have been made have, so far as published work goes, dealt only with the possible relationship of intelligence to ESP ability. Miss Bond's study of retarded children seemed to indicate no correlation. Further work is being carried on now with feeble-minded children. Aside from such problems as whether blindness, age, and sex are correlated negatively or positively with ESP ability, no other questions of correlation have been dealt with even incidentally.

There are under way investigations of possible correlation between results of ESP tests and personality estimates based upon questionnaire and other methods of measurement, but none of these have progressed to the point of publication. The first definite lead toward a correlation between personality characteristics and ESP ability is indicated in Shulman's report in this number on the study of psychotic patients. The differences in results obtained with different classifications of mentally disordered subjects may very well indicate an important correlation between certain personality states and the ability to demonstrate extra-sensory perception. While the author himself recognizes that his report is necessarily a preliminary study and that confirmation is important, particularly to cover the possibility of influence of the experimenter himself, there can be no question that Mr. Shulman has opened up a new field of exploration in the ESP research. It may be men-

tioned that plans are already more or less definite for such work to be repeated in at least three other institutions.

In spite of the fact that the most vigorous controversy has prevailed for some time over the question of the occurrence of extra-sensory perception, and although the thought of precognition is still more difficult, nevertheless the publication of the first report of experiments on the precognition hypothesis has been followed by a curious and practically unbroken silence. Is it possible that the pre-shuffling card-calling work reported in the March number of the JOURNAL is just too incredible for comment, or is it more likely that commentators have caught the feeling which pervades the experimental group itself; namely, one of watchfully awaiting the outcome of the further explorations now under way?

This number of the JOURNAL introduces the research stimulated by the results reported in the March number, results which gave evidence of the ability of subjects apparently to predict the order of a pack of cards as it would be after the experimenter shuffled it a given number of times. The question raised by such pre-shuffle calling was whether the experimenter might not be unwittingly using ESP ability to guide him in advantageously rearranging the cards in the shuffling process so as to bear out the predictions to a sufficiently extra-chance degree to give the deviations reported. The present report deals with the experiments conducted to find out whether it is possible for subjects thus to use ESP ability in the shuffling of cards so as to cause them to match to an extra-chance degree other packs or equivalent series of symbols which are screened from sensory contact. The reader must judge for himself whether or not this work, with its evidence of an "ESP shuffle," completely undermines the case tentatively made for the precognition hypothesis.

For many years there has been recognized by some of the workers in the field of extra-sensory perception the importance of obtaining if possible a mechanical instrument combining the features of selection, presentation, and recording of the stimulus object in the ESP tests, as well as recording the choice and even perhaps the number of hits made by the subject. The problems of engineering, as well as the cost, have retarded most investigators from progressing far toward this desired goal. Most successful has been Mr. G. N. M. Tyrrell; and Tyrrell has been asked to describe his apparatus for the readers of this JOURNAL, partly with the hope that its presentation will stimulate the production of other machines, perhaps with modifications and adaptations. There

are probably many critics of the research who would be relieved to find the whole process put on a mechanical basis. If such an instrument can be devised and introduced without interfering with the occurrence of the phenomenon to be measured as Tyrrell's reports would indicate has been accomplished in his case, it will undoubtedly have served a purpose.

* * *

For the large majority of readers of this JOURNAL the pronouncement by the President of the Institute of Mathematical Statistics published in the December number was doubtless a sufficient ruling on the adequacy of the mathematics used in the evaluation of results. However, many problems still remain involving certain refinements and certain adaptation of statistical method to the problems constantly confronting investigators in this research field. This has led to the mammoth investigation herein reported by Greenwood, of the results of matching cards against subjects' calls for which they were not intended, to the extent of a half million single matchings—the largest empirical chance control series on record in the history of mathematical statistics. Against this huge background of data, Greenwood proposes to test one by one the various hypotheses that are in use or are to be used in the evaluation of the results of ESP tests. Mathematics knows no surer way of checking these hypotheses, and Greenwood's findings will give final answer on a number of points. Most significant is the point already evolved in this first of a series of studies based on this material; namely, that the method of measurement actually in use is the closest approximation that could have been obtained from available systems.

In Rhine's monograph on extra-sensory perception an empirical formula for combining critical ratios from a number of separate experiments was suggested. In this number, Bubb reports the derivation of a formula to replace Rhine's empirical one. While results from the use of either are similar the new formula is more exact than the old.

* * *

The research in extra-sensory perception has been ably reviewed on many occasions by the most competent scientific writers in America, but an unusually pleasing addition to this history was the recent publication in the Spring number of *The American Scholar* of two interpretative articles of the highest excellence. ESP was viewed from the psychological side by Professor Gardner Murphy of Columbia in "Dr. Rhine and the Mind's Eye," and from the mathematical side by Professor E. V. Huntington of Harvard, in "Is it Chance or ESP"? Pro-

fessor Huntington's closing question, "If mathematics has successfully disposed of the hypothesis of chance what has psychology to say about the hypothesis of ESP?" is answered in one way by Dr. Murphy's final remark: "The leverage for the replacement of 17th-century naive mechanism by other conceptions more characteristic of 20th-century scientific adventure is, I think, research on extra-sensory perception." These articles should be "required reading" for students (and critics) of ESP.

THE EXPERIMENTER-SUBJECT RELATIONSHIP IN TESTS FOR *ESP*[1]

J. G. PRATT and MARGARET M. PRICE

ABSTRACT: The writers found that in independent research projects under closely parallel conditions one obtained highly significant scores while the other found only chance results. A logical consideration of the possible causes for this discrepancy indicated that it was likely due to a difference in the experimenters' approach to and handling of subjects. A joint experiment was made to determine if this were the case.

In this experiment, the investigator who formerly had gotten significant results chiefly created the social situation for the subject while the investigator who previously had obtained chance scores closely supervised the testing conditions, checking, and recording. In Section I of the experiment, some subjects were handled in what was thought to be a "favorable" manner while others were treated "unfavorably." Under these exploratory attempts to get a contrast of two ways of handling subjects in the same series of tests, only chance results were obtained. In Section II, the investigator handling the subjects adopted the approach which had previously given significant results. Under these conditions significant scores were obtained at about the same average as in the successful experimenter's independent work (5.53 as against 5.52).

The writers conclude that failure to find evidence of ESP in card tests may be due to an unfavorable experimenter-subject relationship. An effort is made to characterize the difference in approach of these two investigators.

INTRODUCTION

WHILE efforts to repeat the experiments in extra-sensory perception reported in 1934 by Rhine (5) have resulted predominantly in confirmation[2] of his results, several investigators have conducted card-guessing tests without finding any evidence of ESP. Only five failures to confirm have been published (1, 3, 6, 8, 9); two have been presented at psychological meetings, and others have been mentioned in newspaper articles.

Any research leading to conclusions so unusual as ESP must naturally pass through a stage of verification by independent investigators.

[1] Presented at the thirty-third annual meeting of the Southern Society for Philosophy and Psychology, April 15-16, 1938, at Knoxville, Tenn.

[2] These confirmatory studies all either appeared originally or were critically examined in this JOURNAL.

The fact that some of the attempts to repeat these experiments have led to confirmation of the original results while other efforts in the same direction have failed suggests that the conditions favorable to finding the phenomenon have not been sufficiently understood and described. It will be recognized that some failures to demonstrate a phenomenon occurring against a background of more numerous successes cannot in themselves be taken as a disproof of its occurrence. On the other hand, such discrepancies in results cannot occur without cause, and it is of the greatest importance in the understanding of ESP to know to what they are due.

As important as the problem raised by the failures is, there are peculiar difficulties which may discourage its experimental solution. Some of these difficulties are rooted in the general ignorance of what affects ESP, how it functions. The greatest practical difficulty, however, is the fact that the people who are successful and those who fail at finding evidence of ESP are almost always independent investigators in separate laboratories and are thereby handicapped in determining experimentally the causes of the discrepancies in their results.

In the present research, an opportunity was afforded to compare two experimenters' results of ESP tests made under similar conditions and to investigate the possible reasons for a contrast found. One investigator made a series of tests for ESP with only chance results at the same time that another, in the same laboratory, in a closely parallel series, was finding highly significant extra-chance scores. This situation offered an opportunity to focus attention more closely on the role of the experimenter in the ESP tests.

BACKGROUND OF THE EXPERIMENT

The writers, both members of the Parapsychology Laboratory staff at Duke University and both experienced in ESP research, began, in the fall of 1937, independent research projects in Oxford Orphanage, Oxford, N. C.[3] The procedures which the two experimenters used were similar in what appeared to be all essential respects. M. M. P. tested boys and J. G. P. girls. The tests used by M. M. P. were those she had used previously in a study of blind children (4). Her experimental project was the comparison of ESP scores of seeing and nonseeing children.[4] J. G. P. was interested in making a general study of ESP ability under various conditions of testing. Each investigator

[3] We wish to express our deep appreciation for the splendid cooperation received from the members of this Institution.

[4] The results of this comparison will be reported in a later number of the JOURNAL.

used a variety of ESP tests in which sensory perception was carefully ruled out by one of the following precautions:

1. The cards to be called were kept screened from the subject's sight and out of reach.

2. The cards were wrapped in paper and then sealed in heavy manila envelopes.

3. The key cards for Blind Matching[5] tests were arranged, out of the subject's sight, face down in five shallow boxes, and then covered over with blank cards.

The investigators, during the period in which they worked independently (from October 12 to February 1) each tested 40 subjects ranging in age from 8 to 16. Usually the work was confined to one afternoon and evening of each week. The children were tested individually or in pairs for about a half-hour at a time. After each day's work each investigator arranged to test some subjects further and to have new ones in the places of those dropped. New subjects were selected at random by the teachers and supervisors in the institution.

The 40 subjects tested by M. M. P. made a total of 29,325 trials which showed a positive deviation from mean chance expectation of 609 hits, and an average of 5.52 correct per 25 trials. The critical ratio of this result is 8.6, which clearly indicates that the average is reliably different from chance.

Meanwhile J. G. P. had witnessed 53,975 trials. His 40 subjects scored only 81 more hits than the mean chance expectation, an average of 5.04 per 25 cards. The standard deviation of this number of trials is 95, so the results are clearly no more than could be expected by chance. This was a clear-cut case of failure to find evidence of ESP in an extensive series of tests.

The discrepancy in the results of the two experimenters forcibly raises the question of the reason for the contrast found, especially in the light of the great similarity of the working conditions. It is important to consider respects in which the two series were not exactly parallel to see whether they offer any explanation of the facts.

1. The difference in the sex of the two groups of subjects has already been mentioned.

2. A difference in the number of trials in the two series is quite marked. This is largely due to the fact that J. G. P. tested for four more days than M. M. P. However, the possible significance of the speed and distribution of trials will be discussed further in the light of later work observed jointly.

[5] See Glossary.

3. There were slight differences in the test procedures used.

4. M. M. P.'s scores may have been due to an inadequacy of the testing and recording conditions in her experiments to eliminate other hypotheses. (As J. G. P.'s results are not reliably different from chance, this question need only be raised for those of M. M. P.)

5. The possibility remains that the discrepancy arose from a difference induced in the subjects as a result of the way in which each experimenter handled his subjects.

It was impossible to decide conclusively on purely logical grounds which one or more of these five differences in the two series really were important in producing the contrast in results. However, from the available experimental facts, it seemed likely that the difference was largely, if not entirely, traceable to the last factor, that of the experimenter-subject relationship. Previous studies have shown no reliable difference between the sexes in ESP scores. The differences in procedures have not produced comparable differences in results hitherto. In the Parapsychology Laboratory at Duke University, where these investigations have been more extensive than elsewhere, it has long been known that two experimenters might obtain different results even when both use the same tests and the same subjects. Rhine (5) and Stuart and Pratt (7) have suggested that the manner in which the subject is approached might have an all-important effect on the results of the tests. Clark and Sharp (2) reported that the same subjects got such widely different results with different experimenters as to suggest that the attitude and actions of the experimenter may influence a subject's success. However, it appears advisable to have a direct experimental attack on this problem. The present paper is the first step toward an understanding of the personal influence of the experimenter in the ESP tests. In attacking this general problem, this experiment deals only with the logically prior question of whether a failure to obtain evidence of ESP in a series of tests may be due to an unfavorable experimenter-subject relationship. If a personal role of the experimenter is found, attempts to define experimentally the favorable and unfavorable ways of handling subjects would then be in order.

PROCEDURE

An experiment was planned to test the hypothesis that the discrepancy in the earlier results of the two investigators was due to differences induced in the subject by the personal approach to and handling of the subjects. The procedure was designed to eliminate the other possible causes which were listed above.

1. An approximately equal number of boys and girls were used as subjects.

2. The number of trials per day was varied throughout the experimental period.

3. The possible effect upon the scores of slight differences in test procedures was controlled by using two tests, one chosen by each experimenter from among those he previously used.

4. Inadequacy in the manner of testing and of checking and recording the results were guarded against in two ways. Both experimenters were present at all the tests and each observed the precautions taken at each step. Furthermore, J. G. P. was relieved of having to deal personally with the subjects and could give his undivided attention to the precautions for eliminating sensory cues and for the accuracy of the checking and recording of the scores.

The testing techniques used were as follows:

Blind Matching: In test I, a set of cardboard receptacles just large enough to receive the cards and ½ inch deep was used for the five key cards. The cardboard receptacles were attached in a row to a strip of heavy cardboard. In arranging the key cards for a run, the key-card rack was taken completely out of the subject's sight behind an opaque screen and there the five key cards were slipped face down, one into each of the five boxes, without the faces being seen by the experimenter. A blank cover card was then placed over each key card. The concealed key cards were placed before the subject, the subject was handed a pack of shuffled ESP cards, and he proceeded to sort these, face down, into five piles opposite the five boxes so as to have as many cards as possible match the key cards. Experimentation was facilitated by using two key-card racks. While M. M. P. was putting the subject through one run, J. G. P. took the second rack behind the screen and arranged the key cards for the next. A different set of key cards was used for each run during any period of testing a subject. At the end of each run the score was obtained by the experimenters, who turned the key cards out face up upon the table and then sorted the "hits" in each pile from the "misses." In addition to counting the hits as they were sorted, both experimenters re-counted and the score was recorded immediately.

Pack Matching: Test II consisted of using two packs of cards, one of which the experimenter shuffled behind a screen out of the subject's sight, and the other of which the subject then arranged at his leisure in any way he wanted so as to make the order of cards in his pack correspond as closely as possible to the order of those in the experimenter's pack. When the subject felt that his arranging of the cards

had reached a favorable point, he handed his pack to the other experimenter, and the two investigators proceeded to compare the two packs card for card. Both experimenters, as well as the subject, observed this process of comparison; the number of hits were counted aloud, and the score recorded at once at the end of the run.

The experimental period began on February 8 and lasted for seven days through March 22. The first five of these days comprise Section I of the experiment. During this time an effort was made to distinguish experimentally between a "favorable" experimenter-subject relationship and an "unfavorable" one. The "favorable" condition consisted in taking time with each subject for preliminary conversation, encouraging the subject while the tests were in progress, and generally trying to create a pleasant atmosphere for the work. The "unfavorable" condition was an attempt to obtain a control result in a parallel series; in this, the subjects were put immediately to work without any conversation, were not encouraged during the tests, and were "slighted" socially in the test situation. The control subjects were excluded from conversation as much as possible without offending them. The difference between the "unfavorable" and "favorable" conditions will be discussed further in a later section. In Section II, consisting of the last two days' testing, this procedure was discarded because of the restraint that it placed upon the experimenter; instead, for this period M. M. P. took with all the subjects the approach which she had used and found to be successful in her previous work.

RESULTS

During Section I of the experiment, the subjects made 15,625 guesses of ESP cards. These trials yielded a positive deviation from mean chance expectation of 64 hits, an average of 5.10 per 25 trials. The "favorable" condition consisted of 5,225 trials with a positive deviation of 55; the "unfavorable," 10,400 trials with a positive deviation of 9. The results of Section I were not significant, either as a whole or by conditions.

Section II consisted of 3,650 trials with a positive deviation of 79, an average of 5.53 hits per 25 trials. These two days' results show the significant critical ratio of 3.24, with odds against chance of 1,667 to 1. During this period, when M. M. P. worked freely in the manner to which she was accustomed when working alone, the average was almost exactly that which she obtained alone (5.53 as against 5.52).

During the entire experimental period a total of 35 subjects were tested, 17 boys and 18 girls. In general, they did equally well: one

boy and one girl obtained individually significant scores (a critical ratio of 2.5 or more). The differences in the period of working independently therefore cannot be attributed to the sex of the subjects.

A comparison of the two techniques shows that 11,750 trials were made at Blind Matching with a positive deviation of 44 hits; and 6,325 trials were made at Pack Matching with a positive deviation of 97. In addition there were 1,200 trials under miscellaneous testing conditions, all with the cards completely screened from the subject and out of reach. These had a positive deviation of 2.

DISCUSSION

The results of Section II, which are reliably different from chance, were obtained under conditions which rule out all explanations except that of ESP. Sensory cues could not affect the scores under the conditions of the tests (see pages 88-89). All of the work and the recording were witnessed by two experienced experimenters as well as by the subjects. As one of the investigators alone in a much longer series had obtained only chance results, the hypothesis of error can have no bearing here. Whether the hypothesis of chance is eliminated completely depends, of course, on whether the odds of 1,667 to 1 are considered adequate. It is to be observed again, however, that the average in Section II exactly parallels that which M. M. P. obtained when working alone.

The actual extent to which the results bear upon the question of the experimenter-subject relationship depends upon a more detailed analysis of the experimental period. On the first day the first 5 subjects made a total of 2,675 guesses with an average of 4.99 per 25. M. M. P. kept the fifth subject for only about 10 minutes when she sent her away and a conversation very much like the following took place between the investigators:

M. M. P.: "What's the matter with these subjects of yours?"

J. G. P.: "I don't see anything wrong with them."

M. M. P.: "Well, they don't talk! They are as silent as clams! How are you going to know what to do if they don't talk?"

J. G. P.: "Well, they guess cards. Isn't that enough?"

M. M. P. "No. You can't test their ESP if you can't get them to open up. Now when the next subject comes in we are not going to bother about working at all until we get the subject opened up. We are simply going to talk to her."

The next subject came, the entire half hour was spent in lively general conversation, and no testing was done until the next subject

due arrived at his specified time. Then both were tested together, the conversation being continued throughout the period. The subject who had been initiated into the situation by conversation averaged 6.9 for seven runs, the other 5.9. Both subjects were then dismissed.

Another subject came, she was engaged in conversation without working for one-half hour, then was tested with the next to arrive. The subject who was "initiated" scored an average of 6.4 for 10 runs, and the other 4.9.

Due to the suggestion arising from this result, namely, that it might be possible to test the effect of the experimenter-subject relationship by having a single investigator adopt different attitudes toward each of the two subjects present at the same test, M. M. P. was urged to continue this practice of "slighting" half the subjects and fraternizing with the others. The "favorable" condition previously mentioned consisted of introducing a subject to the situation by one-half hour of general conversation before starting the tests and then continuing the conversation during the test. The "unfavorable" one was characterized by starting to test a subject without delay and in deliberately keeping him out of the conversation as much as possible. Two subjects, one in each of the two conditions, were usually tested together in Section I. M. M. P. used these two objectively-defined conditions against her expressed better judgment for the next four experimental days. She insisted all the while that she did not feel capable of "slighting" subjects and claimed that she was not able, under the conditions imposed, to establish with any of the subjects the experimenter-subject relationship which she considered most favorable. This effect upon the experimenter of deliberately trying to make the conditions unfavorable for some subjects probably accounts for the low averages obtained during the first days of joint experimenting. The results of Section I showed that she was not able to get a contrast in results in a single session in which depressing conditions were used with half the subjects. Instead, all the subjects got approximately mean chance scores, the "favored" subjects averaging only slightly better than the others.

After the period of work on March 5, M. M. P. stated that if she were allowed to work without any restrictions upon handling subjects but were allowed to get the best scores she could from all of them, she felt confident of obtaining better results. Section II of the experiment was distinctly foreseen as a test of the importance of the experimenter-subject relationship under conditions most nearly approximating those under which M. M. P. had worked independently.

While the results of Section I may reasonably be attributed to

chance, they become indicative of the importance of the experimenter-subject relationship in ESP tests when considered in relation to the entire series. During Section I M. M. P. was attempting to adapt her approach to comply with specified rules. This involved, as we have seen, attempts to achieve at the same time favoring and depressing conditions. The restrictions placed upon the investigator by the rules not only led to her expressed protest of working under handicap, but actually gave the only extensive series of trials made by this experimenter which have not had a significant average. This indicates that the favorable relationship may be (and in the case of this experimenter *must* be) a spontaneous one which is disturbed when the investigator is placed under certain formal restrictions.

Objectively, the clearest difference between the two experimenters detectable from their records is that of the number of trials. M. M. P. generally obtains about two-thirds as many scores as J. G. P. in an equal time. It might appear superficially that this indicates a difference in the rate of guessing. Actually, however, the subjects worked for each experimenter at their own natural rates, and the difference in the total number of trials was due to a difference in the amount of time taken between runs. M. M. P. avoids hurrying her subjects and encourages a free social atmosphere in which general conversation flourishes. She has a way of directing the subject's attention to the task at hand and of explaining what is to be done with a few simple, well-directed remarks. After she has once got the subject started at the test, she usually manages to get his attention on something other than the testing process. As long as the responses are made without hesitation or evidence of too serious effort, she permits, and even encourages, the subject to talk about something beside the test. When she sees any evidence of a misunderstanding of the test, she corrects the subject as briefly as possible, sometimes in a scolding tone, and then returns at once to the interrupted conversation. In this way her subjects are first consciously oriented with respect to their task and are then diverted as much as possible to permit them still to keep that orientation. They "play" at the tests without getting any strong feeling of needing to win or of having a duty to perform.

J. G. P.'s approach, on the other hand, has always been to direct the subject's attention as much as possible to the task and to keep it there. He felt that full explanations were due the subject, even when the information given was not necessary for taking the tests. All-in-all, his subjects were given a more serious view of the tests, looked upon them as a pleasant duty, and were encouraged to concentrate upon the

task at hand to the exclusion of other things, e.g., conversation while working.

This brief description of a successful experimenter-subject relationship is offered without any claim to finality, for the favorable and unfavorable experimenter-subject relationships must themselves be experimentally defined, and this is yet to be done. There is no indication at present that an approach which is successful for one investigator will be successful for others. As desirable as such generalizations would be, they cannot be made without a broader basis of facts.

The present report goes far to establish an effect of the experimenter-subject relationship on the results of tests for ESP. This investigation focuses the problem raised by ESP experiments with chance results and points the way to an experimental solution of the problem. The results suggest forcibly that failures to find evidence of ESP may be indicative of an unfavorable approach to and handling of subjects. The writers hope that other investigators encountering this problem will be able to conduct further experiments along these lines.

SUMMARY

Two experienced investigators, both favorable to ESP research, conducted tests under very similar conditions with children in an orphanage. After a period of working independently, a comparison of their results showed that one had obtained scores indicative of ESP while the other had not. The conditions were similar enough to suggest strongly that the experimenter-subject relationship was the important factor, but they were not similar enough to prove it. A further period of joint experimenting was planned to test this hypothesis. The experimental procedure was designed to control other possible explanations of the earlier discrepancy of results. This series, observed throughout by both investigators, was divided into two periods. In Section I, an effort was made to handle some subjects in a "favorable" manner and others in an "unfavorable" one. For this section the results were not significant. In Section II, the experimenter who had previously been successful handled the subjects in the manner which was natural to her, while the second investigator controlled all the tests and recorded the results. The results of Section II were reliably different from chance and had the same average as the first experimenter had obtained while working alone.

The present experiment focussed a problem which bears upon the interpretation of any failure to find evidence of ESP in card guessing tests. In view of the fact that the Duke work has already been inde-

pendently confirmed in other laboratories many times, the failure of any investigator personally to find subjects with ESP ability may be due to an unfavorable experimenter-subject relationship in his tests.

REFERENCES

1. BAKER, K. H. Report of a minor investigation of extra-sensory perception. *J. Exp. Psychol.*, 1937, **21**, 120-125.
2. CLARK, C. C. AND SHARP, V. Group tests for extra-sensory perception. *J. Parapsychol.*, 1937, **1**, 123-142.
3. COX, W. S. An experiment on extra-sensory perception. *J. Exp. Psychol.*, 1936, **19**, 429-436.
4. PRICE, M. M. AND PEGRAM, M. H. Extra-sensory perception among the blind. *J. Parapsychol.*, **1**, 143-155.
5. RHINE, J. B. Extra-sensory perception. Boston: Bruce Humphries, 1934.
6. SOAL, S. G. A repetition of Dr. J. B. Rhine's work in extra-sensory perception. *Jour. Soc. Psych. Res.*, London, 1937, **30**, 55-58.
7. STUART, C. E. AND PRATT, J. G. A handbook for testing extra-sensory perception. New York: Farrar and Rinehart, 1937.
8. WILLOUGHBY, R. R. Prerequisites for a clairvoyance hypothesis. *J. Appl. Psychol.*, 1935, **19**, 543-550.
9. WILLOUGHBY, R. R. Further card-guessing experiments. *J. Gen. Psychol.*, 1937, **17**, 3-13.

A STUDY OF CARD GUESSING IN PSYCHOTIC SUBJECTS

ROBERT SHULMAN

ABSTRACT: Screened matching tests for ESP were given to 141 patients in a mental hospital. The patients were grouped into 14 clinical classifications. Of these, one group, the manic-depressive depressed, gave significant positive scores. Another group, those with involutional melancholia, gave consistent, although not significant, negative scores.

Mr. Shulman conducted the work as a research project in the Psychology Department at Bard College. The experiments were carried out in the Hudson River State Hospital, Poughkeepsie, New York.

EXTRA-SENSORY perception ("ESP"), defined by Rhine as "response to an external event (perception) not presented to the known senses," has been and is being investigated in college students, children, blind persons, and other groups, under a variety of conditions. While the writer was working with psychotics, studying symbolism and abstractions, it became evident that they were suitable subjects for studies, using one of the card guessing techniques developed by investigators of ESP. In these individuals psychomotor and emotional processes deviate widely from the normal and provide suitable conditions for studying their effect on the ESP scores. It is difficult if not impossible to find such exaggerated states in normal subjects. Therefore it was found desirable and feasible to carry out the research reported in this paper.

Two important questions arise in connection with this study. First, are psychotic subjects capable of scoring significantly above chance expectation when tested by one of the methods developed for investigating ESP ability, the STM (screened touch matching) technique introduced by Pratt (3). Second, how do the different classifications of psychotic subjects according to hospital diagnosis compare on the test?

The experimental work of this study was begun at the Hudson River State Hospital, Poughkeepsie, New York, in June and concluded in November, 1937. Due acknowledgment is made to the superin-

tendent, Dr. Ralph P. Folsom, to the clinical director, Dr. John Y. Notkin, and to those physicians on whose services the work was carried out, for their invaluable assistance, encouragement, and cooperation.[1]

Subjects: The primary limiting factor in the selection of subjects for the study was that they were male patients in a state hospital for the insane. Two hundred and fifty subjects from fourteen different classifications (see Table I) were used in this research. As a pre-

TABLE I

SUMMARY OF DATA BY GROUPS

Groups—based on hospital diagnosis of condition	No. in group	Total No. of Trials	Total No. Hits	Average per 25	Total Dev.	Stan. Dev.	Critical Ratio
1. Dementia Praecox—Hebephrenic	11	5,625	1,141	5.06	16	30.00	.53
2. Dementia Praecox—Paranoid	7	3,500	706	5.04	6	23.66	.25
3. Dementia Praecox—Catatonic	5	2,250	443	4.95	— 7	18.97	— .37
4. Alcoholic Psychosis	10	4,750	934	4.93	—16	27.56	— .58
5. Paranoid Condition	6	3,500	698	4.99	— 2	23.66	— .08
6. Psychoneurosis	6	3,000	602	5.02	2	21.91	.09
7. Psychosis with Psychopathic Personality	7	3,000	570	4.76	—30	21.91	—1.37
8. General Paresis	25	14,000	2,824	5.04	24	47.32	.51
9. Manic-Depressive—Manic	9	4,500	896	4.98	— 4	26.83	— .14
10. Manic-Depressive—Depressed	12	12,250	2,600	5.38	150	44.27	3.388
11. Involutional Melancholia	9	7,250	1,383	4.72	—67	34.06	—1.937
12. Psychosis with Epilepsy	14	6,750	1,394	5.16	44	32.86	1.34
13. Psychosis with Cerebral Arteriosclerosis	12	5,250	1,003	4.78	—47	28.98	—1.62
14. Psychosis with Mental Deficiency	8	3,875	761	4.89	—14	24.90	— .56
Totals	141	79,500	15,955	5.05	55	112.78	.48

liminary step to working on each of the services, the physician-in-charge, who was most familiar with the condition of the patients at the time, was asked to prepare a list including all of the patients whom he judged might cooperate effectively in the test situation. An attempt

[1] The writer wishes to express his appreciation to Dr. J. B. Rhine for his criticisms and suggestions throughout the work, to Dr. C. R. Carpenter of Bard College for his assistance in the preparation of this paper, and to Dr. Gardner Murphy of Columbia University for his suggestions and criticisms.

was made to test all patients listed by the physicians. Those who refused to try the test or who worked only if continually urged, as well as those who showed little or no understanding of the card guessing procedure, were eliminated from the tests after one or two runs. Of the 250 male patients with whom tests were attempted, 141 were found to be sufficiently cooperative and to show a clear enough understanding to warrant continuation of the test in their case beyond the first or second run of 25 cards. The scores made by all of the subjects who completed 10 runs or more (141 patients) are included in the data. Thus a good sampling of the hospital's cooperative male patients was secured.

The work was done on five separate services, each of which included the patients of from three to eight wards. Since it was imperative, however, to secure as many cooperative patients as possible, about half of the total time of the study was spent on the North Wing service, which had all of the groups represented and whose patients were sufficiently well-adjusted to permit of their helping with simple types of manual labor.

Materials: ESP cards,[2] in standard packs of 25 cards composed of five each of the five different symbols, were used throughout the experiment. A plywood screen, two feet long by eight inches high, was mounted on side supports which raised the bottom of the screen two inches above the table. Although the tables used varied from service to service as to size, there was an average distance of three feet between the subject and the experimenter.

Methods: The STM (screened touch matching) method was used throughout the experiment. The subject was seated at a small table across from the experimenter and between them there was a screen placed in the middle of the table. The five key cards (each marked with a different one of the regular five ESP symbols) against which the cards of the shuffled deck were to be matched, were placed face up; one in each of five open cardboard boxes, directly underneath the bottom of the screen. Thus the subject could point to the key cards but he could not see any of the table on the experimenter's side of the screen when seated. Although there was a two inch aperture between the bottom of the screen and the table to allow the experimenter to see the pointer, the subject's head was at all times in full view of the experimenter. This fact made it possible to check any attempt on the part of the subject to bend down and look at the cards. As a further precaution, the cards were shuffled and held by the experimenter below

[2] See frontispiece, JOURNAL OF PARAPSYCHOLOGY, Vol. 1, No. 1, March, 1937.

the level of the table until called, the table serving as an additional screen, so to speak. The cards were held face down and placed face down by the experimenter on the table as each guess was made.

A simple explanation of the test procedure was made to each subject. He was shown that the pack of 25 cards consisted of five cards of each of the five symbols. The object of the test was to match the cards in the deck with the proper key cards. Each guess would be made by pointing to the key card which he felt matched the top card of the deck. This top card would then be placed directly behind the box containing the key card indicated by the pointer and another guess would then be made. After all of the 25 cards had been guessed, they would be checked to see how many had been matched with the correct key card. No attempt was made to explain the purpose of the test, but rather to secure the subject's interest in the "guessing game." Many of the patients, particularly when starting the tests, seemed eager to "get a high score" and in some cases appeared to enjoy the "game" as a welcome change from the daily routine. The cards were checked after each run of 25 cards and the subject was told his score. The tests were given in sets of ten runs daily, with at least one day elapsing between the first and second set of ten runs made by each subject.

The advantage of this technique is that it does not require the subject to learn the symbols and it permits him to work at his own speed and rhythm of pointing or guessing. Furthermore, the method apparently interested psychotic subjects and yet conditions were completely controlled by the experimenter. Under the described conditions, it was impossible for the subjects to see the cards at any time before or after calling them. However, the procedure was not so stereotyped or monotonous as to discourage the subject. The experimenter sought to secure good rapport with the subject and to create a strong interest and good cooperation. A number of the tests were witnessed by physicians, nurses, and attendants at the hospital and a group of tests (mostly of manic-depressive depressed, paretic, and epileptic patients) was witnessed by Dr. Harold R. Phalen of Bard College. Most of the tests were carried out with the subject and the experimenter alone in the room.

The hospital diagnosis was taken as the criterion for grouping the subjects. It should be noted, however, that the subjects within a group varied from extreme excitement or deeply depressed states to apparently quite normal behavior, when tested. Individual subjects also showed great variation in activity and "mood" from one testing period to the next.

Results: 1. Occurrence of ESP ability among psychotic subjects: The results of this study indicate that some psychotic subjects are capable of scoring significantly above chance expectation when tested by one of the methods delevoped for investigating ESP ability (the STM technique). Six in all, of the 141 subjects tested for at least ten runs, or 4.25%, scored significantly above chance expectation in a series of ten runs or more (see Table II). Of these six, one subject succeeded in maintaining a just significant score, critical ratio 2.5, when he had been given a total of 40 runs or 1,000 trials. The scores of the other five subjects were not significant when all of the runs that they had done were totaled for each individual.

A total of 79,500 trials was given to 141 subjects, resulting in a positive deviation of 55 hits from the number expected by chance. This score gives a critical ratio of .48 for the entire study (see Table I). It is therefore not a significant departure from the mean chance expectation. A significant difference was found, however, between the scores of certain groups, as reported in the following section.

2. Comparison of scores made by groups: Table I gives a summary of the group scores. A study of this table shows that the manic-depressive depressed group scored much higher than any of the other groups. It is the only group having a significant critical ratio (3.4). Of the 12 subjects comprising this group, ten scored above chance expectation, although not significantly, and the other two scored only slightly below chance expectation. A study of individual and group scores indicates that the significant critical ratio of the group is not due to the high scores of any one individual but to the consistent scoring of the majority of the group above chance expectation. This consistent trend within the group is one of the most significant findings of the study. Three of the subjects in this group scored significantly in a series of ten runs or more. The individual scores of the subjects making up the group are given in Table III.

In contrast to the manic-depressive depressed group, the involutional melancholia group scores give a critical ratio (—1.94) well below chance expectation, although it is not a statistically significant negative deviation. There are nine subjects in this group, of whom eight scored below chance expectation and one just slightly above. Here, too, as in the manic-depressive depressed group, the critical ratio of the group is not due to the low scoring of any one individual but to the consistent trend of scoring for the entire group below chance expectation, with one exception. The individual subject scores of the involutional melancholia group are given in Table IV.

TABLE II

SUBJECTS SCORING SIGNIFICANTLY ON 10 RUNS OR MORE

Subject	Classification	Runs	Hits
1. M. C.....	Manic-Depressive, Depressed...	10	66*
		10	37
		10	52
		10	47
		10	41
		Total. 50	243
2. H. H.....	Manic-Depressive, Depressed...	10	69*
		10	58
		10	39
		10	55
		10	61
		10	36
		10	42
		10	56
		Total. 80	416
3. E. S.....	Manic-Depressive, Depressed...	10	39
		10	61
		10	68*
		10	55
		10	48
		10	48
		10	51
		Total. 70	370
4. W. D.....	General Paresis.....	10	62
		10	67*
		10	41
		10	48
		10	40
		10	54
		Total. 60	312
5. W. B.....	Psychosis with Cerebral Arteriosclerosis.....	10	48
		10	67*
		Total. 20	115
6. F. B.....	Dementia Praecox, Hebephrenic.	10	57
		10	64
		10	54
		10	57
		Total. 40	232*

*Significant score.

TABLE III

MANIC-DEPRESSIVE DEPRESSED GROUP; SUMMARY OF INDIVIDUAL SUBJECT SCORES

Subject	Trials	Hits	Ave. per 25	Dev.	S.D.	C.R.
1. M. C.	1,250	243	4.86	— 7	14.14	— .49
2. H. F.	1,000	219	5.475	19	12.65	1.50
3. H. H.	2,000	416	5.20	16	17.88	.90
4. A. K.	1,000	212	5.30	12	12.65	1.50
5. H. S.	1,000	195	4.875	— 5	12.63	— .39
6. T. S.	1,000	222	5.55	22	12.65	1.73
7. J. T.	1,000	210	5.25	10	12.65	.79
8. E. S.	1,750	370	5.28	20	16.73	1.19
9. A. M.	500	118	5.90	18	8.94	2.01
10. I. S.	500	113	5.65	13	8.94	1.45
11. D. A.	750	168	5.60	18	10.95	1.64
12. O. W.	500	114	5.70	14	8.94	1.56
Totals	12,250	2,600	5.386	150	44.27	3.388

A significant critical ratio of 4.40 is found for the difference between the total scores of the manic-depressive depressed and the involutional melancholia groups. The critical ratios of two other groups —psychosis with psychopathic personality (1.37) and psychosis with cerebral arteriosclerosis (1.62)—were well below chance expectation but not significantly so. One other group, psychosis with epilepsy, had a critical ratio of 1.34; relatively high for this study. The critical ratios of the remaining nine groups do not vary more than .58 above or below chance expectation.

Discussion: This must be considered a preliminary, exploratory study and similar but more complete investigations on the psychotics in other hospitals should be carried out in order to substantiate and extend the findings of this study. The results however, seem to in-

TABLE IV

INVOLUTIONAL MELANCHOLIA GROUP; SUMMARY OF INDIVIDUAL SUBJECT SCORES

Subject	Trials	Hits	Ave. per 25	Dev.	S.D.	C.R.
1. I. C.	1,000	198	4.95	— 2	12.65	— .15
2. J. M.	1,000	185	4.62	—15	12.65	—1.18
3. T. P.	750	141	4.70	— 9	10.95	— .82
4. G. S.	1,000	207	5.14	7	12.65	.55
5. F. T.	1,000	191	4.78	— 9	12.65	— .71
6. N. A.	500	90	4.50	—10	8.94	—1.11
7. J. H.	500	89	4.45	—11	8.94	—1.23
8. J. H3.	500	92	4.60	— 8	8.94	— .89
9. P. D.	1,000	190	4.75	—10	12.65	— .79
Totals	7,250	1,383	4.72	—67	34.06	—1.937

dicate that some psychotics are able to score significantly above chance expectation on the STM method of testing for ESP ability. It is suggested that the scores of these subjects are effected by frequent shifts in degree of attention. It was noted repeatedly that subjects who had shown a strong interest during their early runs lost interest and became distracted after 8 or 10 runs. Many individuals returned to the next testing period in a much more depressed or excited state, therefore a more distracted one, than had previously been the case. Accompanying this loss of interest, there was apparently a drop in scoring ability in many cases. However, no definite quantitative statement of these shifts in degree of attention as related to scoring ability can be made at this time, because of the impossibility of measuring the degree of attentional shift and "moods" under the exploratory conditions of this experiment. The statement must remain a "clinical judgment" until more detailed work has been done on the problem of correlating changes in various psychological states with scoring ability on ESP tests.

The most suggestive results of this study are found in the significant difference between the total scores of the manic-depressive depressed and the involutional melancholia groups. These results demand further analysis and discussion. Apparently shifts in attention occurred as frequently in the subjects of these two groups as in others. Nevertheless, there is evident a consistency in the scores of each of these two groups which suggests that something other than a pure chance factor was operative in determining the scores. What possible hypothesis can be suggested to explain this consistency? Before attempting to answer this question, it is necessary first to consider and compare the accepted clinical pictures of these two psychoses.

White (5, p. 152) states that "the depressed phase of the manic-depressive psychosis manifests itself by three cardinal symptoms, each one of which is opposed to the corresponding symptom of the manic stage, viz., 1. Difficulty of thinking. 2. Psychomotor retardation. 3. Emotional depression." Each of these symptoms may manifest itself with any degree of severity and the three symptoms may not all vary proportionately. The mildest grade of depression, which is the one concerned in this study, is called *simple retardation.* As a rule, the depressed subjects in this experiment varied from apparently normal behavior to this grade of mild depression. In these cases all responses were slower than normal. Fisher (1, p. 277) states that in the manic-depressive psychosis "the patient responds slowly and without interest to questions; his writing is slow; apparently all his mental processes

are retarded." In general, the process of association of ideas is quite retarded, but in spite of slowness of apprehension and thought, *consciousness and orientation are well retained.* Probably one of the most distinguishing characteristics of the manic-depressive psychosis is the fact that the behavior of these patients differs from the normal essentially not in kind but in degree, hence the difference is not a qualitative but rather a quantitative one.

Melancholia is essentially a disease of the period of involution, rarely occurring before fifty in men. The most common form of this psychosis is called *agitated melancholia.* In these cases, the picture is one of an agitated depression, *with consciousness usually clouded and some disorientation for time, places, and persons.* The milder cases of involutional melancholia sometimes, according to White (5, p. 244), resemble the depressed stage of the manic-depressive psychosis as to psychomotor activity, with slow movements and slow responses to questions in a low voice. Repetition of one or two phrases and a stereotyped motor reaction are quite common, particularly in the agitated phase of the psychosis. Several writers also comment on the increased egoism and negativism in the melancholic, with the development of nihilistic ideas.

The fundamental differences between these psychoses are found in the interpretation of them. White (5, p. 159) considers the manic-depressive psychosis as a type of *extraversion reaction.* "The manic-depressive psychosis is not to be characterized so much by the nature of the difficulty (conflict) with which the patient has to deal as by the way in which he deals with it. This method I have described," writes White, "as a 'flight into reality' which is the characteristic of the manic phase, while the failure to deal adequately with the difficulty is manifested by the depression of the depressed phase." In the case of the involutional melancholic, however, interest in the outer world is given up or gradually diminished and there is evident a concentration of mental energy on some particular idea or object, with an accompanying increase in negativistic tendencies.

There is a considerable difference of opinion among various writers as to the correct classification of involutional melancholia as a distinct type of psychosis. However, let the following three assumptions be made:

1. Negativism is present to a significantly greater degree in the patients of the involutional melancholia group than in the patients classified as manic-depressive depressed.

2. Voluntary control of ESP ability is possible. See Rhine (4, p. 115) and Pegram (2).
3. Negativism will affect ESP ability by producing scores having a negative deviation; i.e., if the subject is instructed to try for a high score, his negativistic tendencies operate to cause him to make a below chance score.

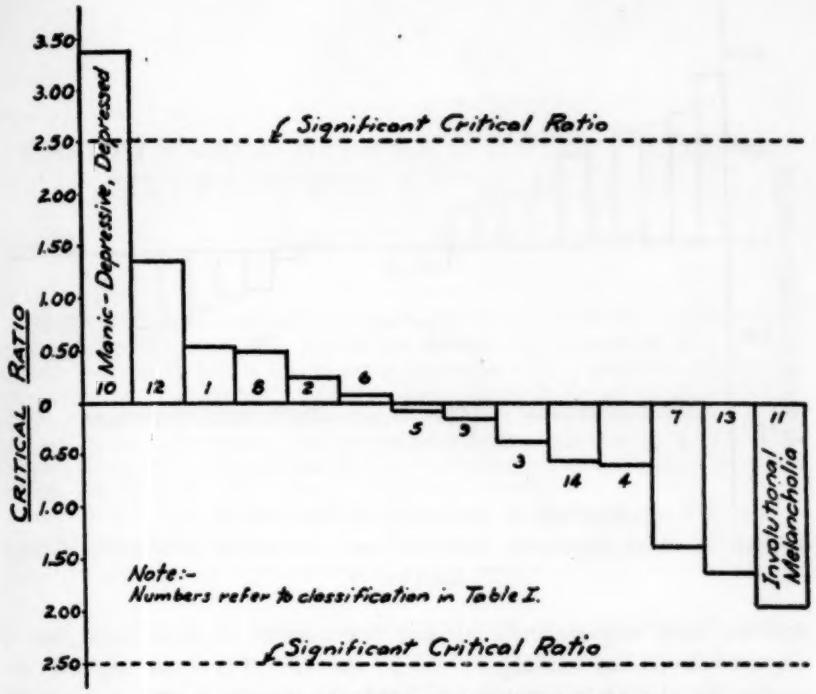
Then, it appears logical to suggest that it is possible to account for the negative deviation of the involutional melancholia group scores on the basis of the negativistic tendencies of the patients in this group.

It also appears probable that an important difference between the two psychoses as applied to this study centers around the degree of attention and perception of external events found in each group. Although the manic-depressive depressed subjects show a general slowness in reacting to stimuli, they are not "shut-off" from external interests. The melancholics, on the other hand, tend to be more subjective and egocentric, thus being distracted and dissociated from the task of card guessing as well as from other external stimuli. It should be noted that several of the melancholic subjects kept up a monotonous repetition of irrelevant phrases during the tests, although when they were questioned directly, they appeared to understand that they were trying to match the cards in the deck with the proper key cards.

The manic-depressive depressed subject more nearly approaches the normal with regard to the factors of external interest and apparently, lack of negativistic tendencies. As a group, they showed a more consistently applied interest to and consciousness of the task at hand than did the melancholics. The subjects of the involutional group showed an apparent "blocking" of external interest. Therefore, in the light of the preceding discussion, it seems permissible to *suggest* that a clue to the reasons for the significant difference between these two groups as they scored in this study may be found in the difference in degree of effective attention, external interest, inversion of "thinking and feeling," and negativistic reactions.

Editorial note: A research similar to that reported above by Mr. Shulman was begun by Mr. Seymour Van Wiemokly, a psychology graduate of Duke University, using as subjects for the tests patients in the New Jersey State Hospital at Morristown. Mr. Van Wiemokly's study was interrupted for a time by his medical studies but will later be resumed and a full report published.

A note is permitted here, however, to say that using a method comparable to that of Mr. Shulman, and ***knowing nothing of the scores***



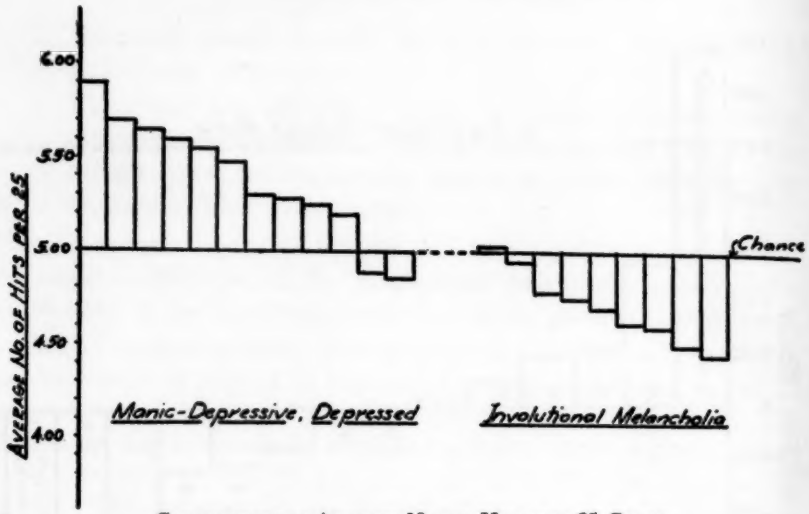
Comparison of All Groups by Critical Ratio
GRAPH I.

obtained by Shulman, Van Wiemokly obtained averages in the three classifications tested (before interruption) that are strikingly close to Shulman's. For these three classes of patients the results are as follows:

Classification	Experimenter	No. Sub.	Runs	Av. /25
Paranoid schizophrenia	Shulman	7	140	5.04
	Van Wiemokly	4	100	5.01
Epileptic psychosis	Shulman	14	270	5.16
	Van Wiemokly	5	133	5.13
Involutional melancholia	Shulman	9	290	4.72
	Van Wiemokly	4	100	4.77

It may be seen that the appropriate averages for the two independent investigators differ at most by .05 hits per 25 trials; whereas the differences between classes is .12, .24, and .36 for Van Wiemokly and .12, .32, and .44 for Shulman.

Another point that stands out is that for both experimenters the lowest average was given by the involutional melancholia patients and



COMPARISON OF AVERAGE NO. OF HITS PER 25 CALLS
Subjects of Manic-Depressive, Depressed and Involutional Melancholia Groups.
GRAPH II.

that for each experimenter all but one patient in this class gave a negative deviation, totaling 11 out of the 13. The total negative deviation for the 13 involutional melancholia patients in 390 runs is 90, which is 2.23 times the standard deviation.

These data of Van Wiemokly reinforce the suggestive character of Shulman's results.

REFERENCES

1. FISHER, V. E. An introduction to abnormal psychology. New York: The Macmillan Co., 1929.
2. PEGRAM, M. H. Some psychological relations of extra-sensory perception. *J. Parapsychol.*, 1937, **1**, 191-205.
3. PRATT, J. G. Clairvoyant blind matching. *J. Parapsychol.*, 1937, **1**, 10-17.
4. RHINE, J. B. New frontiers of the mind. New York: Farrar and Rhinehart, 1937.
5. WHITE, W. A. Outlines of psychiatry, 12th ed. Washington: Nervous and Mental Disease Publishing Co., 1929.

THE TYRRELL APPARATUS FOR TESTING EXTRA-SENSORY PERCEPTION

G. N. M. TYRRELL

ABSTRACT: As an alternative to shuffled cards to provide the random choices necessary for ESP tests, Mr. Tyrrell has developed a mechanical device which presents the subject with a chance series of choices with a one-in-five probability for correctness. The subject's task is to choose in which one of five boxes before him an electric lamp is lit. The correct box may be determined directly by keys operated by the experimenter, the pattern fixed by the keys may be shifted in an unknown order by a commutator, or the boxes may be determined in a wholly random and unknown pattern by a mechanical selector.

Mr. Tyrrell's research was reviewed by Burke M. Smith in the first number of this Journal for 1937.

ALTHOUGH packs of cards provide the most obvious way of testing [the faculty of] Extra-Sensory Perception in a quantitative manner, wherein the choices are so circumscribed as to be amenable to mathematical treatment, the use of more complicated apparatus offers certain substantial advantages. The apparatus described below has been already reported on in the *Proceedings* of the English Society for Psychical Research,[1] but the technical details were given only in outline. It is with pleasure therefore that I take this opportunity to describe its main features in greater detail. It may be as well first of all briefly to summarise the advantages and disadvantages of what may perhaps be called the mechanical line of approach to the extra-sensory problem.

Advantages. As compared with the method of using cards the machine herein described has the following advantages. (1) It affords a means of experimenting under conditions of which the subject is normally unaware. (2) It affords a means of changing from a condition known to the subject to one unknown or from one unknown condition to another, without the subject being aware that any change of condition has been made. (3) It allows of tests being carried out with great rapidity. (4) It allows of the use of a motor response. This is not necessarily an advantage for all subjects, but is an advantage where

[1] *Proc. S. P. R.* Vol. XLIV, p. 99.

aphasia appears connected with the condition of the subject under test, as is not infrequently the case. (5) An inexhaustible supply of randomly distributed numbers, as amongst the five boxes concerned, is given by the Mechanical Selector with no trouble at all to the experimenter. (6) The result of each experiment is automatically and permanently recorded. (7) There are no loopholes for fraud, modes of misuse being automatically guarded against. (8) Where precognitive experiments are concerned, it is possible to deal with very short intervals of time between the event and the response by measuring the overlap of lines on the recording tape. These short intervals could not well be measured with a stop-watch. (9) The operator can be entirely eliminated if desired.

Disadvantages. (1) The apparatus is much more expensive than cards. (2) Though not very large, it is not portable in the sense in which packs of cards are portable. (3) There is the theoretical objection that the mechanism may go wrong. On the other hand, testing is very quick and easy.

*General description.** The apparatus is fastened onto two tables, each measuring 30 inches by 17 inches. On the subject's table is a row of five wooden boxes, each 1¼ inches wide, with sloping lids overlapping the boxes in front and firmly held down by rubber bands at the sides. Each box contains an electric pea-lamp, which can be lit by the operator. See Photos 1 and 2 (screen removed). The subject's table is connected to the operator's table by flexible wires and can be moved to any part of the room. It is provided with a screen the full width of the table (30 inches) and 5 feet high from the floor, and is so placed that the subject is completely hidden behind the screen. Under the edge of the table is an electric sounder.

The operator's table is placed in the corner of the room so that the operator sits with his back to one wall and his right hand to another, the table being in front of him. The back and left hand of the table are screened with a wooden screen 5 feet high from the floor, an opening about 2 feet wide being left for entry. Thus the operator and his apparatus is doubly screened from the subject, who usually sits 6 or 8 feet away. The general lay-out of the apparatus which faces the operator may be seen from Photo 3. Beginning from the left, the double switch changes from the use of keys to the use of mechanical selector. This has now been converted into a foot-switch. Behind it, the top of the Delay-action Relay† stands up. The smaller

* The whole apparatus can be worked off 26 dry cells.

† This device permits selection of a given box by the machine with delay in lighting until the subject's choice is made; see description below.

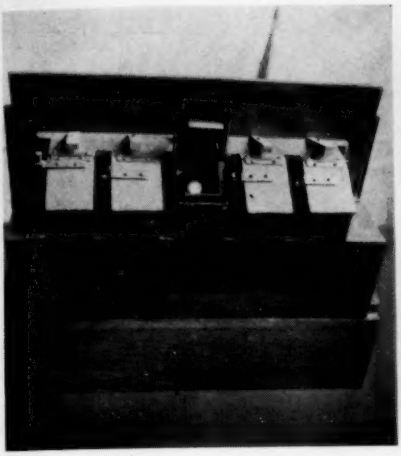


FIG. 1. Boxes from front

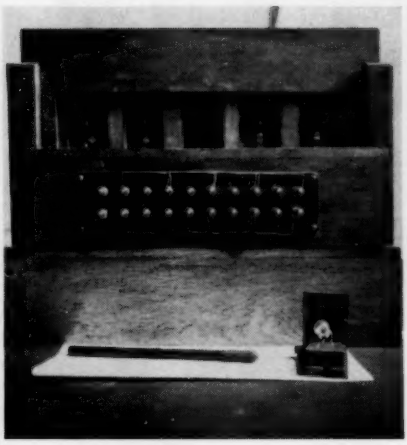


FIG. 2. Boxes from rear

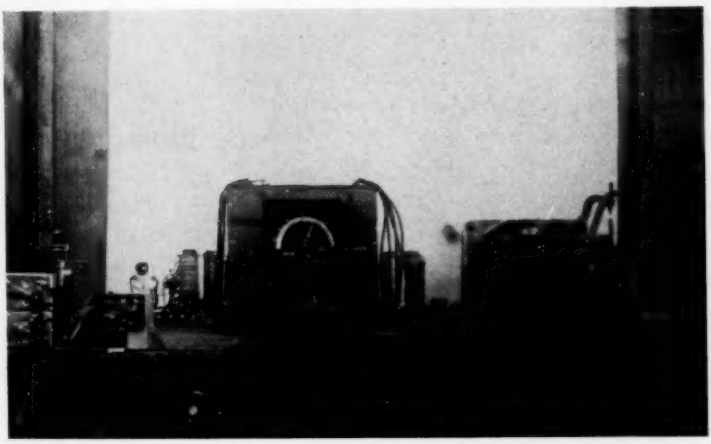


FIG. 3. Operator's table

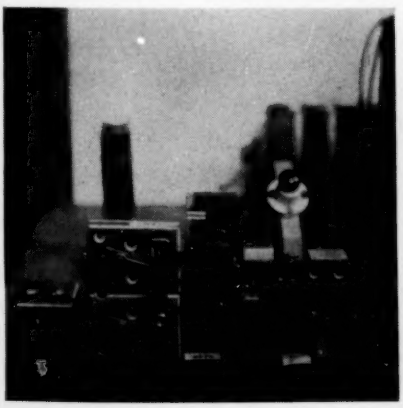


FIG. 4. Switching arrangements, Signal lever, Counter, Synchronizing lamp

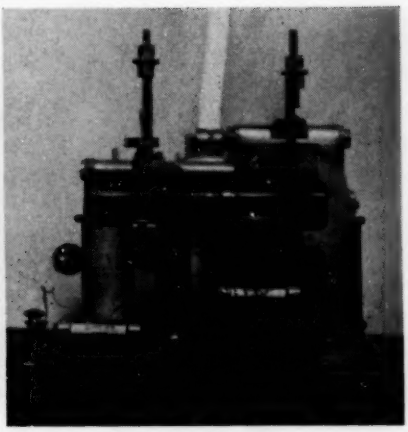


FIG. 5. Recorders

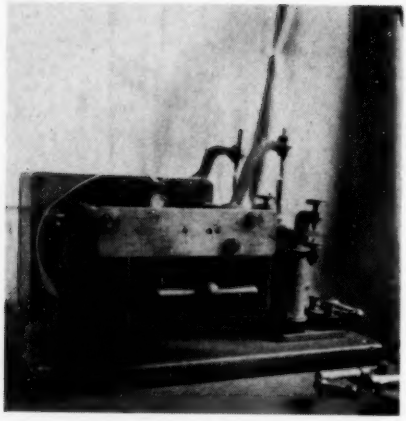


FIG. 6. Recorders

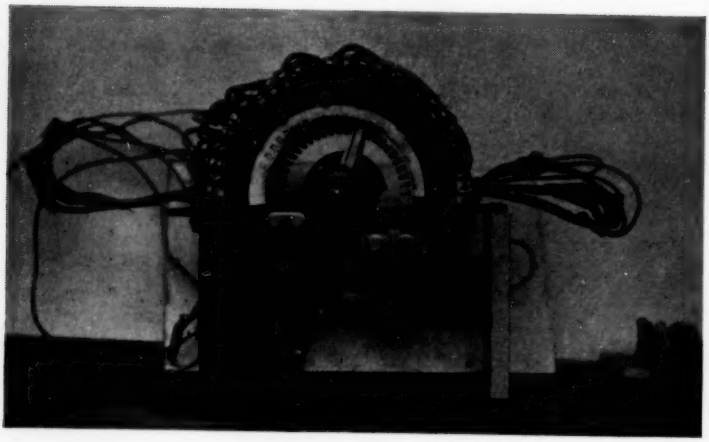


FIG. 7. Commutator

switch a little to the right is for cutting out the delay-action when required. Above it is the synchronising lamp, and a little to the left the lever which works the Trial-counter and the sounder on the subject's table. These will be seen better in Photo 4. In the middle, standing up in its box, is the Commutator with the door of the box open. The bell-push causes the commutator to revolve and gives various transpositions of the wires running to the lamps. Behind the commutator is the row of relays for working the Success-recorder (cover removed). Next, the five keys can be seen, their contacts dipping into cups of mercury. Immediately to the right again is the key which operates the mechanical selector. Behind this is the double recorder, which records Trials and Successes on the same strip of moving tape, driven by clockwork. The end-view of this is seen in Photo 5, and another view on Photo 6. The mechanical selector is placed in an adjoining room so as to be inaudible. Its construction is similar to that of the commutator, which is shown in Photo 7. The operator's relays are carefully silenced and in addition are placed under a thick oak cover faced with felt and screwed to the table.

Operator's Choices. The operator has at his command several conditions. He can (1) use the five keys, selecting their order as he goes along. (2) Use the five keys, taking their order from a list of random numbers. (3) Use the mechanical selector by pressing and releasing the key on the right. (4) Allow the subject to work the mechanical selector himself by the act of opening and closing the boxes. (5) Use the delay-action. (6) He can change from any one arrangement to any other in the middle of a series by moving the silent foot-switch. (6) He can perform any experiment with known or unknown settings of the commutator.

Method of Carrying out an Experiment. Subject and operator sit in their places as described, out of sight of one another. The operator starts the tape running and says the word "Ready." He then presses one of the five keys (or the selector key) with his right hand and at the same time raises the lever with his left. This causes the sounder to click on the subject's table and simultaneously records a trial on the counter. The subject, hearing the click, raises one of the box-lids. This works the trial-recorder, making a dash on the tape (it is a double dash if he has opened the box with the light in it) and the trial-recorder lights the synchronising lamp. Thus the operator holds down his key during the brief interval during which the synchronising lamp is alight, meanwhile bringing the lever down again with his left hand. When the lamp goes out the same process is repeated and so on for a hundred

trials. During this time the operator does not speak, but at the end of the series he says the word "Stop." The recorder is stopped and the tape torn off with the full record on it. The hundred trials will normally have occupied about a minute and a half or a little more. The experiment is thus simplicity itself. (An interesting thing to notice is that one inevitably works these experiments in a sort of rhythm. One starts by pressing the key and giving the signal to the subject *before* the latter responds. But very soon the click of the sounder on the subject's table and the click of the trial-recorder which is worked by the subject's opening the box, occur *together*. The subject must be selecting the box *before* the lamp is actually alight. A series of slow trials was made, in which the lamp was lit every time at least half a second before the signal to open a box was given, but this did not give any increase in the score.)

The Boxes. These are very carefully made. The lids are revetted and sunk back under the hinge and the facings lined with fine velvet. When closed therefore there is a complete light trap. They are held down with rubber bands, one on each side. The contacts, at first inside the boxes, were afterwards placed outside and worked by arms projecting from the lids. The lamps were run rather dull. Raising the lid $\frac{1}{16}$ of an inch in front is sufficient to close the contact and cause a record of the choice to be made.

The Recorder. Photographs 3, 5, and 6 show this. There are two ink-wheels, side by side, and each pair of magnets, when energized, raises one of them against the tape. One records each trial and the other each success. Diagram 1 shows how this is done. $P_{1\text{-}5}$ are

AUTOMATIC SUCCESS RECORDER

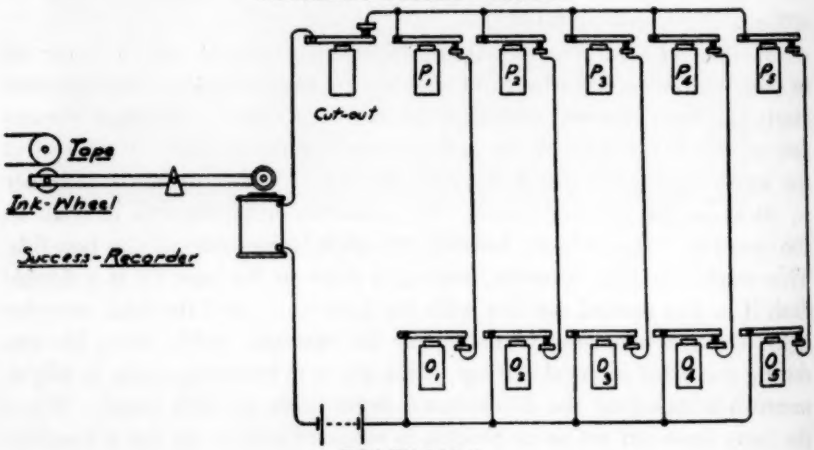


DIAGRAM 1

the percipient's relays each energized when its respective box-lid is opened. O_{1-5} are the operator's relays, each energized when its respective lamp is lit. The diagram shows how the success-recorder will only work when any corresponding *pair* of relays, such as P_1 and O_1 are working together. It shows also the cut-out, which breaks the circuit of the success-recorder if more than one box-lid is opened at a time. This is seen better in Diagram 4. The cut-out is set so as not to work when the current of one lamp is passing through it, but to work if the current of two or more lamps is passing. Thus if more than one box is opened at a time the apparatus can register nothing but failures.

The Commutator. This is shown in its case in Photo 3 and out of its case in Photo 7. The mechanism is the same as that of the Selector, which is described below, the only difference being that in the case of the commutator there are five pairs of rotating arms and five semicircles of 25 contacts over which the arms pass. Each contact seen in the photograph is, so to speak, five deep. The wires from the five keys are connected to the five rotating arms. The wires from the five lamps are connected to the five rows of contacts at one end and are then criss-crossed among the groups of five all the way round the semicircle. Thus, wherever the arms stop they connect keys to lamps in a different order. Certain of these positions are however straightforward one-to-one connections. When the door of the case is shut and the push-button pressed the arms rotate rapidly and it is impossible to know on what contacts they have stopped.

This arrangement not only cuts out the possibility of telepathy. It also cuts out any possibility of normal communication from the operator to the subject, as by unconscious whispering, etc. In addition to this, in cases where the operator is making his own selection of the sequence of keys pressed, it breaks up any possibility of coincidence of number-habits between himself and the subject, whether these are simple number preferences (such, for example, as a preference for the middle number) or sequence-preferences, such as a frequent repetition of the sequence 1, 2, 3. Each setting of the commutator will distribute such preferences differently among the lamps. The way in which the Commutator is connected into the circuit is shown in Diagram 4.

The Selector. This consists of identically the same kind of instrument as forms the commutator, except that four of the five pairs of arms are removed. Photo 7 therefore illustrates it. Diagram 2 shows the exact way in which the 25 contacts are connected. An irregular

distribution of five contacts of each number is connected together and in this way any slight differences in the resistance of different contacts that might exist is evened out and distributed. The contacts themselves are projecting teeth and the arms contain split ends which hold the teeth between the two sides. The motor operates a ratchet mechanism which moves the arm on one contact at a time. It may be said that this mechanism has been found far superior to those using an ordinary electric motor.

SELECTOR

Details of Connections

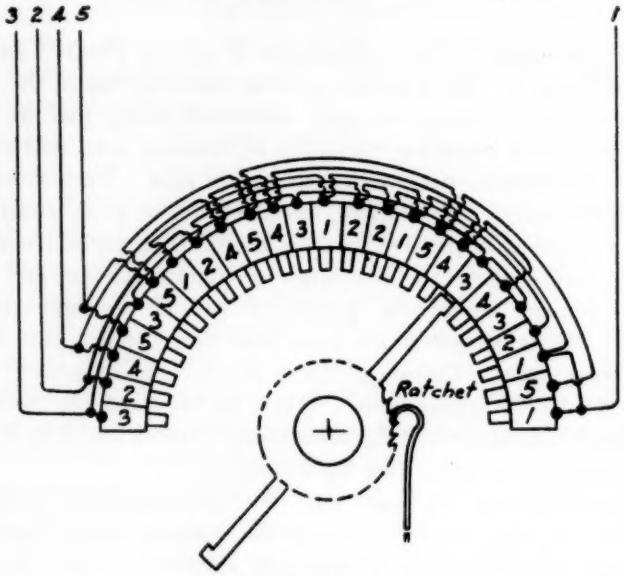


DIAGRAM 2

The speed of working is high, a half-revolution of the arms occupying 0.43 second, so that the time taken in moving from one contact to the next is 0.017 second. It is absolutely impossible therefore for the operator to manipulate the key so as to make the arm stop on a predetermined contact. To test this I took the selector out of its case and placed it on the table in front of me. I then fixed my eye on the contact on which I wished the arm to stop and tried to operate the key so as to make it do so. I failed 9 times out of 10 experiments. With the selector in its case it is of course utterly impossible to influence it in any way.

Samples of 2000 trials have been taken from time to time to test the random nature of the sequence of numbers given by the selector.

A mathematical analysis of one such sample was published in the *Journal* of the English S.P.R., vol. XXX, p. 83. It shows that the numbers are impartial, but even if they had not been, the use of the commutator through which the wires pass on their way to the lamps (see Diagram 4) would render such impartiality harmless, a fact which the author of the above analysis does not seem to have realized. It must be remembered that it is only *coincidence* between similar departures from true randomness on the part of selector and subject which can affect the results.

Automatic Use of the Selector. One very interesting use to which the present apparatus can be put is that of eliminating the operator altogether and causing the subject to work the whole thing himself. Diagram 4 shows how this is done. An automatic relay is introduced into the circuit at AB. P_{1-5} are the percipient's relays. B_{1-5} are the box-lid contacts. L_{1-5} are the lamps in the boxes. O_{1-5} are the operator's relays. When any box-lid is opened the current passes through the automatic relay and energizes it. This causes two things to happen. (1) The contact at A is broken, which stops the selector from rotating. (2) The contact at B is closed, which completes the lamp-circuit and allows the lamp chosen by the selector to light. In Diagram 4, for example, we may suppose that the subject has opened No. 2 box, which closes the contact B_2. This will have stopped the relay on one particular contact, which, in the diagram, will have lit the lamp L_2. The subject will have scored a success, and the two relays, P_2 and O_2 will have operated simultaneously and will have brought the success-recorder into action (see Diagram 1). But if the selector chances to have stopped on any other contact, the trial will have been a failure. It will be seen that any success scored with this arrangement will be success in Precognition, since the subject must know where the selector is *going* to stop when he chooses his box. But it is precognition in which voluntary human action plays a causal part, since it is his own action which causes the selector to stop. Any theory that it is inference from contemporary clairvoyance has to take into account that the arm is moving so fast that it is difficult for the eye to follow it, and is covering 25 contacts in less than half a second. That the contacts are connected together in a most complicated way; that the selector is connected to the commutator by several yards of twisted flexible wires; that the commutator itself is full of complicated cross-connections, all of which must be traced out before it can be predicted what contact will light a particular lamp.

The contacts are adjusted so that A breaks before B closes. Thus

there is no sparking on the contacts of the selector for no current passes through them unless the arm is at rest.

When using the selector in the ordinary way, a double-ended key is substituted for the automatic relay. A is the back-stop contact and B the front. Precognition with this key is tested by getting the subject to open the box before the selector is stopped.

DELAY-ACTION DEVICE

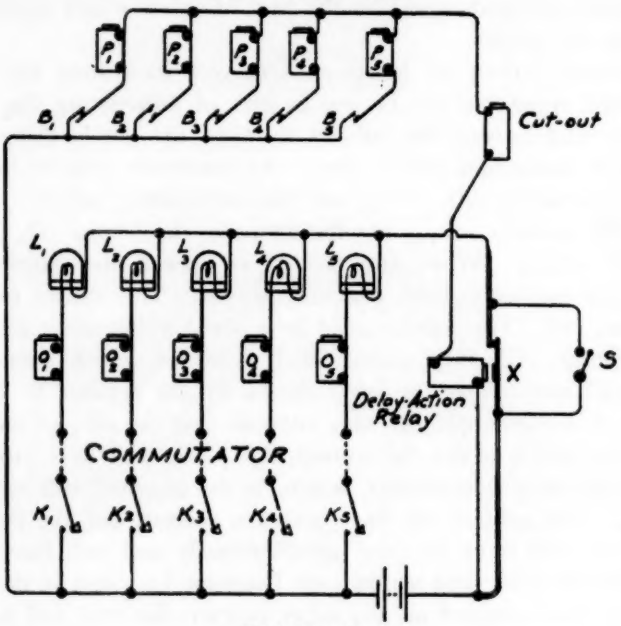


DIAGRAM 3

Delay-action Device. Diagram 3 illustrates this. It provides another way of testing precognition and also tests the theory of hyperaesthesia of the normal senses. K_{1-5} are the silent mercury keys. The other letters are as before. The return-wire from the box-contacts is taken through a simple relay, which, when energized, closes a contact X. This contact is placed in the common return-wire from the lamps.

Now, when any key is pressed by the operator, the lamp which will be lit by it is selected; but the lamp does not light because the relay-contact at X is open. But as soon as the subject opens any box-lid it closes one of the contacts B_{1-5}. This energizes the delay-action relay, closes the contact X and lights the selected lamp. Success with this arrangement apparently needs precognition since, owing to the commutator, the operator does not know what lamp he has selected. It

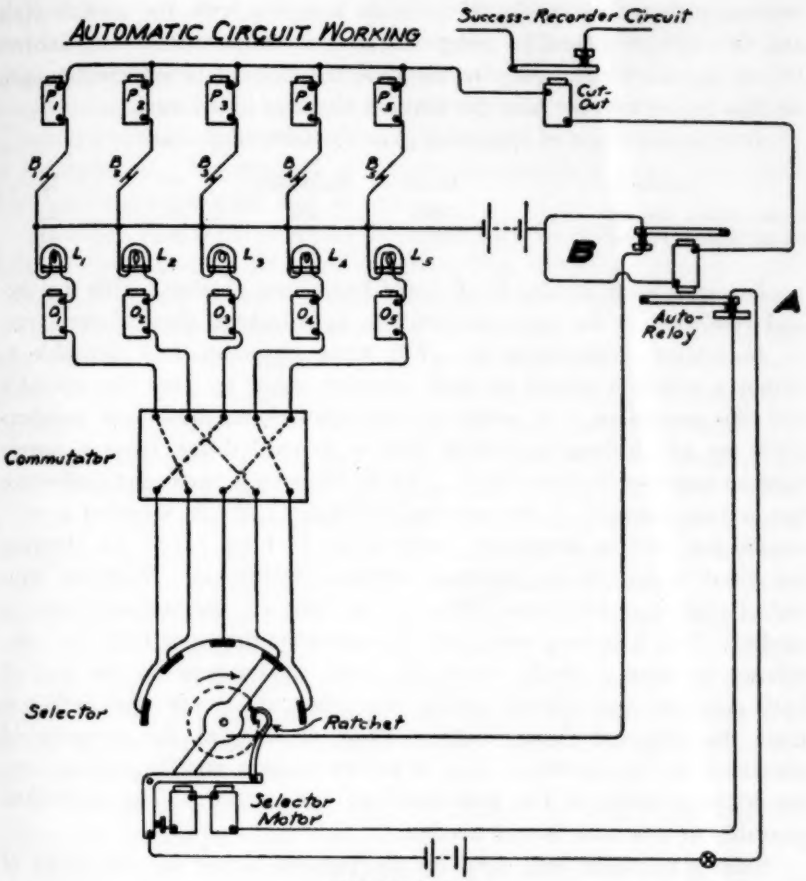


DIAGRAM 4

might be said that the subject clairvoyantly notes which key has been pressed and clairvoyantly traces out the connections through the commutator to the lamp. But, as has been said, the pressing of the key and the opening of the box are in practice simultaneous, so that this explanation does not sound plausible. Any rise of electric potential of the selected lamp will be shared by all the others on account of the common wire.

Success with this arrangement also cuts out all possibility of normal help by supposed leakage of light from the lamps past the box-lids, radiation of heat or any other kind passing through the lids etc. since the lamp can give none of these indications when it is not alight. Closing the switch S throws the arrangement out of action. It is only

necessary therefore to do 50 trials in a series with the switch open and 50 with it closed in order to have a comparison. The subject knows no difference, since by the time the box-lid is sufficiently open for him to see into the box, the lamp is alight in either case.

Intermixed trials of this kind gave the following results:

Condition.	Trials.	Successes.	%	X.
Delay-action cut out.	845	242	28.7	6.3
Delay-action in use.	855	224	26.2	4.5

No significant results at all could have been obtained with the second condition if the hyperaesthetic or light-leakage theory were true.

Individual Trial Recorder. For some purposes it is desirable to obtain a separate record of each selection made by both the operator and the percipient. A series of free choices amongst five numbers made by any human individual will in general differ from a purely random sequence in three ways. (1) In showing a particular preference for certain numbers at the expense of others. (2) In showing a preference for certain sequences, such as 3, 5, 4 etc. (3) In showing too great regularity in intervals between repetitions. With an individual trial-recorder these habits on the part of the subjects may be studied. But it is very important to remember that it is only the *coincidence* between a similar deviation from randomness on the part of *both* operator and subject which can effect the score and deflect it from the expected chance value. Thus, so long as the sequence of selections on the operator's side is known to be a strictly random one, the idiosyncrasies of the percipient do not matter. The individual recorder in this case is not needed.

Such a recorder was used by the present writer in one series of experiments, the arrangement being as follows. Operator and subject sat in separate rooms, each having five keys before them. The usual sounder and synchronising lamp were used. The recorder comprised ten pens drawing ten parallel lines on a wide, moving strip of paper. Each key, when pressed, operated a magnet which deflected its own particular pen and made a peak in the line. Every key pressed by the operator as well as by the subject made its record. Sometimes the operator used a list of random numbers taken previously from the mechanical selector and sometimes he chose the numbers as he went along. The interesting point is that in neither case were any results obtained significantly differing from 20 per cent, although in the cases in which no random numbers were used, opportunity existed for coincidence between operator's and subject's number-habits.

For the study of individual idiosyncrasies, such a recorder is undoubtedly useful, but for the recording of ordinary experiments in which a known random series of numbers is used by the operator, it is unnecessary and also clumsy. It provides an embarrassing wealth of material which is not needed and which it takes a great deal of time to go through. The method of recording described above has been found to be generally the best in practice.

Testing. The apparatus may be tested (1) by setting the commutator to the "straight through" position (i.e. wires not mixed) and working keys and boxes together backwards and forwards from end to end; every trial should then record as a success. (2) By starting keys and boxes from opposite ends, when only the middle number should record as a success. (3) By running through 100 trials, the subject counting each light which appears and comparing with the number of successes on the tape. (4) By unscrewing the brass plugs in the fronts of the boxes and verifying that a lamp lights every time a key is pressed or the selector operated. This may be combined with testing on the recorder. All these tests are very quickly carried out. There is in fact very little risk of getting too *many* successes. Most of the things which might conceivably go wrong would cause a *reduction* in the number of successes. This, too, is true of the effect of the operator and subject getting out of step; but it is impossible for this to happen without its being noticed. Nor can the apparatus be misused. If any box is not quite closed, the recorder draws a long line on the tape and so stops the record.

Of course any mechanical apparatus needs looking after and the above tests should be made frequently.

Another good practical test for the impartiality of the selector is to take 2000 numbers from it and write them down and then open the boxes in the order of these 2000 numbers against the working of the selector. The number of successes thus obtained can be tested for significance. If the selector has any bias this bias will presumably be the same in the second 2000 trials as it was in the first, and will therefore give a deviation from chance-expected value. The commutator must of course be kept in the same position throughout.

Results. So far as Miss G. M. Johnson, the chief subject, was concerned, the use of the selector had become associated with a strong inhibition owing to discouraging failures while the selector was in its experimental stage. This was partly due to a lack of confidence in it, since the experimental selector frequently broke down or needed adjustment, and so spoiled the result. Such fixed inhibitions appear to

be very difficult to eradicate. The subject would score as long as keys were used but would not score when the selector was used. All the evidence pointed to the fact that it was the subject's *belief* about the condition in use which determined the result. If the selector was introduced without the subject's knowledge she would score with it; the moment she became doubtful the score would drop. Neither was it possible to deceive the subject for long since she soon discerned what condition was being used apparently telepathically.

To meet this difficulty, lists of numbers taken from the selector were written down on cards and these were used in connection with the keys. Evidently this condition was equivalent to the use of the selector, since the introduction of the commutator overcame the objection that the subject might have learnt the numbers on the cards; and the subliminal of the subject did not appear to object to it. The following results were obtained with this arrangement.

Condition.	Trials.	Successes.	X.
Numbers from Selector used through the Keys.	7809	1841	7.9

These are clairvoyant results, owing to the use of the commutator, and no possible theory of number-habits can account for them.

General Remarks. Everything has tended to show that the problem of extra-sensory perception is the problem of preparing a route for the externalisation of knowledge possessed by the subliminal self. A certain dreamy state is necessary—a slight departure from the normal state of consciousness—and this state is inhibited by the slightest degree of self-consciousness, discomfort or mental or physical disturbance. Consequently the most important conditions are the *general* ones which affect the contentment, the buoyancy, the confidence and the expectation of success of the subject. It is these to which the experimenter should pay most attention if his work is to be truly scientific, for it is of the essence of scientific work to discover which are the most important conditions for success. Casual experiments with subjects who were strangers have not given encouraging results.

EXPERIMENTS BEARING ON THE PRECOGNITION HYPOTHESIS: II. THE ROLE OF *ESP* IN THE SHUFFLING OF CARDS

J. B. RHINE, BURKE M. SMITH, and J. L. WOODRUFF

ABSTRACT: In a previous article, discussing the results of ESP tests in which the subjects attempted to predict the order of cards as it would be when the pack was shuffled, the writer noted that the significantly high results could not be considered unequivocal evidence for "precognition" until a logical alternative had been tested. This alternative was the possibility that the experimenter might be able to shuffle the cards in such a manner as to "match" the predictive call-series.

In the experiments reported herewith, subjects were given a pack of ESP cards and asked to shuffle them until they "matched" another pack, or a denoted series of 25 symbols. Highly significant positive extra-chance results were obtained from the scores of 203 subjects. When the experiments with screened and sealed cards are summarized the total score from 104 subjects is again highly significant. Tests in which the subject's shuffling was limited in time or to a specified number of acts yielded significant results only in the case of one group. The writers conclude that ESP may assist in the determination of the order of a shuffled deck of cards at least when the subject consciously attempts the task without restriction upon the act of shuffling.

The writers are members of the staff of the Parapsychology Laboratory at Duke University.

IN THE preceding number of this JOURNAL there was reported a lengthy series of experiments which were designed for the investigation of the precognition hypothesis, experiments in which the procedure consisted of a subject's making supposedly predictive calls of the order of a pack of cards as it would be following random rearrangement due to shuffling.

These experiments contributed results which gave positive deviations from mean chance expectation that were significant of the operation of an extra-chance factor. The experimental conditions were such as to prevent the utilization of sensory cues by the person taking the test. The number of experimenters involved, and hence the amount of confirmatory repetition, was large enough to warrant the consideration of the results in connection with the objective; namely, the experimental testing of the hypothesis of precognition.

It was recognized, however, that before these experiments could be regarded as constituting a crucial test of the precognition hypothesis, the shuffling of the cards would have to constitute a *truly random rearrangement.* It will be recalled that in the block of data (113,075 trials) submitted for this consideration, the shuffling of the cards was entirely in the hands of the experimenters. The question was raised, however, whether, if the experimenter himself possessed ESP ability, the shuffling might not be subject to purposive instead of random action through the agency of ESP itself. That is, suppose his capacity for extra-sensory perception of the cards might enable the experimenter to cause the pack shuffled to match to an extra-chance extent the series of predictive calls which had been made by the subject, even though the person shuffling had no sensory contact with either the call series or the faces of the cards. More generally phrased, this involves the problem of determining whether anyone can use ESP ability to arrange cards in a pack by ordinary shuffling techniques.

It remained, then, only to institute tests in which the task of the subject was to shuffle an inverted pack of cards with deliberate intent to cause it to match either another pack of ESP cards or a series of 25 ESP symbols. This paper is a report of all the results of the experiments made to date to investigate this hypothesis.

PROCEDURES AND CONDITIONS

The first general method of investigation consisted simply in shuffling one pack to match another, the order of one or both being unknown to the subject doing the shuffling. The experimenter shuffled one pack of cards, placed it face down upon the table with care to prevent any card face being visible to the subject, and the subject shuffled a second pack until he thought it "matched" that of the experimenter.

In later series, the experimenter's shuffling as well as the pack shuffled by him were *screened from the subject by an opaque screen.* In certain short series the experimenter's pack was sealed in a box and the results checked from a code known only to the experimenter, the checking being done in the absence of the subject.

The experimenter's pack (or series of symbols) is called the target pack (or series). The pack which the subject shuffles with the intent to match the target pack is called the shuffler's pack.

The shuffler's pack is, in the main block of these experiments, shuffled by the subject in his own preferred manner (except for the requirement of the inverted position of the pack) and without restriction upon the time of shuffling or number of shuffles given.

It is recognized that this is not a precise duplication of the pre-shuffle calling experiments reported in the first precognition paper. In those experiments there was either a restricted period of shuffling or a fixed number of shuffles. The objective, however, in this research was to give the ESP shuffle hypothesis the most favorable conditions; and if significant results were obtained under these conditions, to turn, in subsequent investigation, to the more exact duplication of the precognition tests. Some attempts have already been made on this further investigation of restricted shuffling, and the preliminary results will be briefly reviewed below. The problem here, however, was first to test the bare possibility of an influence of ESP upon the shuffling, and the shuffling was therefore unrestricted except by the subject's own choice.

Following the shuffling of the shuffler's pack, the experimenter and subject (when the subject was an adult) laid off the cards of the pack in the order of their occurrence, observing matching or failure to match, and recording the results in numbers of hits (i.e. correct correspondences). When children were investigated, the experimenter himself or the experimenter and an observer—not a subject—handled both packs of cards. When a recorded series of symbols was used instead of the target pack, the record of the shuffler's pack was taken on the standard record sheet and coincidences with the target series checked in the usual manner.

The greater part of the more crucial subdivisions of the data of this report have been independently re-checked for errors. This applies to all of the work in which a sealed target pack was used, and to approximately half of the screened target pack subdivision. Two experimenters (not subjects) were present throughout the experimentation and checking of the data for a substantial portion of the tests which yielded the highest score averages. These considerations will be appropriately specified with the presentation of subdivisions of the data below.

The subjects in the experiments cover a wide range of classification. Children from three institutions (two orphanages and a school for the blind) have been used and have proved to be among the most successful subjects. A considerable number of the subjects were college students, and among the others are several laymen. One of the outstanding subjects was a psychologist, Dr. R. W. George. The number of subjects (not the total of column 2 in Table II because of duplications) is 203.

This research represents the work of 13 experimenters. The part

which each experimenter has played will be indicated in the tables. For the most part, the work has been conducted in the Duke Parapsychology Laboratory, but a portion has been done at other places, primarily at Tarkio College under the direction of Dr. R. W. George and Mr. James MacFarland, of the Psychology Department; in the private investigations of Mr. E. P. Gibson, at the time City Engineer of East Grand Rapids, Michigan; and of Messrs. G. E. Buck and J. L. Michaelson, engineers engaged in industrial work in Schenectady, New York.*

The work reported here has been done at intervals during the period since the autumn of 1934.

RESULTS

General Totals. The main body of this report covers all trials made under the conditions described above, a total of 8,461 pack-matchings, or a total of 211,520 card-matchings. If it be remembered that the report of the first three years of the Duke investigations in extra-sensory perception as reported in 1934 numbered little more than 90,000 card calls, some estimate of the size of this investigation can be obtained. For the purpose of general evaluation, the results of all the tests—without selection or exception—are thrown together in terms of total scores. The total positive deviation of scores for the series of 8,461 pack-matchings is 1,959, which with a standard deviation of 187.7 gives a critical ratio of 10.4. These figures seem to leave no room for doubt that something beyond a random distribution of scores is represented.

Unscreened Target Pack. It will be recalled from the description of the procedures given above that in the later experiments the target pack was screened from the subject while he was shuffling the shuffler's pack. At first, however, the tests were made with the target pack unscreened. It is remotely possible, with the target pack unscreened, that the top card might have had some original or acquired defect by which a subject could, consciously or not, recognize it; this would, of course, have no effect upon scoring unless the subject in shuffling the pack (in the inverted position) was able to detect either tactually or

* The authors of this report wish to emphasize the large measure of credit which is due to the many experimenters who have played a prominent part in the investigation, and to express appreciation to the many subjects who generously gave many hours to the research. The cooperation of the administration and staff of the North Carolina State School for the Blind at Raleigh, of the Oxford Orphanage, at Oxford, N. C., and of the Wright Refuge at Durham, N. C. are gratefully acknowledged.

visually what the cards were in the shuffler's pack as well and was able to place them by a clever legerdemain. This possibility, however improbable it may appear, is sufficient to warrant the emphasis being placed upon the series in which the target pack was screened.

This restriction of emphasis does not, however, make the unscreened work completely devoid of interest and value. The view taken is that the screened work bears out the extra-sensory character of the unscreened by the confirmatory score averages it presents, and lends a certain security to the results. For this reason the data of the unscreened tests are included here.

The total of 136,275 unscreened trials gave a total positive deviation of 1,405, which with a S.D. of 150.6 gives a C.R. of 9.3. This represents an average number of hits of only 5.25 per run of 25, but the maintenance of this average for the lengthy series reported is beyond question an extra-chance phenomenon. Further details of the unscreened work are given in Section A of Table I, where appear also the names of the experimenters who are to be credited with the conduct of the sub-series opposite their names.

The table referred to shows two conditions, namely "high-aim" and "low-aim." Rhine[1] originally, and Pegram[2] more extensively showed that subjects could intentionally call cards incorrectly to an extra-chance degree of success; i.e., when the subject was trying to miss the cards the negative deviation from the mean was significantly great.

In this block of data 251 runs were made under low-aim conditions, with the intent to miss the key pack. The total negative deviation was 25. This is not a significant result. In striking a total of the unscreened work, this deviation is given the positive sign since it occurred in the direction intended and is measurable in precisely the same way as the positive deviation for the normal or high-aim conditions.

A feature of interest in the results of tests made with the unscreened target pack is in their comparison with those in which the target pack was screened or sealed. The averages are 5.25, 5.18, and 5.21 respectively as may be seen in Table I. The differences are not significant.

Screened Target Pack. In 2,728 runs, the target pack was completely screened from the subject's sense organs during the test. The extra-chance nature of the results is evidenced by a critical ratio of 4.63 which is given by the deviation of 494 in the total of 68,200 trials. The odds against such a result occurring by chance alone

[1] *Extra-Sensory Perception,* Boston, 1934, pp. 70 and 80.

[2] "Some Psychological Relations in Extra-Sensory Perception," J. PARAPSYCHOL., 1937, I, 191-205.

are approximately 590,000 to one. In Section B of Table I may be found the presentation of data from the screened target pack series, summarized under the name of the appropriate experimenter, with some information concerning subjects and witnesses.

While the case is clear so far as the exclusion of the chance hypothesis is concerned, interest naturally attaches to the experimental conditions maintained for each series, and to the precautions in the handling of the data. As stated above, in all the 2,728 runs of this section the target pack shuffled by the experimenter was effectively screened from the subject. In addition the shuffler's pack was itself shuffled in an inverted position, even though this could not be regarded as an essential safeguard.

In addition there were 1,789 runs made with screened target packs, two experimenters present throughout, observing all testing, recording, and checking. These gave an average number of hits of 5.20 per 25. They are independently significant but are combined with the sealed work below under totals of double-witnessed trials.

There are two further special features to the screened work section of the data: The first concerns the conditions referred to above as "high-aim" and "low-aim." In most of the runs, 1,774, the subject tried to shuffle the pack to match the target pack; in the remaining 954 runs he shuffled the pack with the intention of missing the target pack and securing as low a score as possible. It will be seen from the table that a negative deviation of 175 was obtained in the 954 runs. This is an average deviation per run of .18; and in the high-aim work the positive deviation also averaged .18. There is no reason to expect a difference in these deviations except on grounds of difference in attitude in the subject and accordingly a close comparison is not surprising. A larger deviation for low-aim than for high-aim was found in the work referred to above by Pegram[3] but was not by Rhine.[4]

As stated above, the same method of evaluation applies to a negative deviation for a low-aim series as applies to a positive deviation in the normal case, and these deviations have been added instead of subtracted for the securing of a general deviation *in the direction intended* for the total screened work as well as for the total of all the work reported.

An additional feature of interest inheres in the work conducted by Mr. MacFarland and Dr. George which consists of 470 runs of high-

[3] *Op. cit.* [4] *Op. cit.*

aim and 929 runs of low-aim tests. All of this work belongs to a larger project which will be reported in full in a forthcoming article. This report is complete for the conditions here concerned.

The methods of MacFarland and George differed only in that call series which had been made by other subjects in other tests were used instead of target packs, and the shuffler's packs were shuffled by the subject to match these call series. The record sheets containing the call series had not been seen by the subject and were kept by the experimenter behind an opaque screen. The checking and scoring have been double-checked. It is interesting to note that this work, which approaches most closely to the case of the earlier precognition report (pre-shuffling calling), is in itself significant (C.R. = 3.7).

Fourteen subjects out of the total number of children tested were either partially or totally blind. The detailed comparison of performance by these blind children with the seeing children of comparable age will be given in a later and more complete report by Miss Price.

Sealed Target Packs. This small subdivision is included for completeness. It consists of 282 runs with a positive deviation of 60. It belongs, in effect, with the screened series, since the sealing of the target pack naturally has the same effect as placing it behind an opaque screen. This portion is not in itself significant, but when merged with the screened series it raises the total to 3,010 runs with a positive deviation of 554 and a critical ratio of 4.94.

This too belongs to a larger project; namely, the use of sealed, opaque envelopes in the ESP tests, a problem which will be the subject of an early report.

Double-Witnessing and Re-Checking. There were 1,789 runs made with screened targets in which there were two experimenters present, and in 272 runs with sealed targets this condition also obtained. This brings the total number of runs with invisible (to the subject) targets plus double-witnessing up to 2,061 and a deviation of 424, with a C.R. of 4.65 and odds of about 600,000 to one. The average is 5.20. See totals in Table II.

Another aspect of interest to careful students of ESP is that of added safeguards in checking. In addition to the double-witnessing mentioned, double-checking or re-checking is possible when the records of both decks are available. In a total of 1,671 runs this independent re-checking has been done. These give a deviation of 348 and a critical ratio of 4.17. The average here is 5.20.

TABLE I

RESULTS OF TESTS WITH UNRESTRICTED SHUFFLE

A. Unscreened ESP Shuffle

Experimenter	Subjects	"High-Aim"			"Low-Aim"			Remarks
		Runs	Dev.	Av/25	Runs	Dev.	Av/25	
Smith, B. M.	15	793	+ 101	5.12				
Pegram, M. H.	67	1383	+ 400*	5.28				
Stuart, C. E.	3	90	+ 38	5.42	12	+ 3	5.25	
Woodruff, J. L.	4	281	+ 73	5.25	223	— 20	4.91	Double witnessed.
Weckesser, M.	2	70	— 5	4.92				
McCartney, L.	14	1151	+ 160	5.13				
Gibson, E. P.	20	830	+ 158*	5.19	16	— 8	4.5	
Buck, G. E. Michaelson, J. L.	4	602	+ 455*	5.75				
Total*	129	5200	+1380*	5.26	251	— 25	4.90	

B. Screened ESP Shuffle

Smith, B. M.	8	306	+ 16	5.05	25	+ 10	5.4	
Smith, B. M. Pratt, J. G.	2	8	— 11	3.62	.....	.....	.....	Double witnessed.
Pratt, J. G. Price, M. M.	23	382	+ 87	5.22	.....	.....	.....	Double witnessed Oxford Orphanage.
Price, M. M.	34	161	+ 122*	5.75	.....	.....	.....	Raleigh School for Blind, Oxford Orphanage, and Duke University.
Pegram, M. H.	1	2	+ 1	5.5				
Gibson, E. P.	1	8	0	5.00				
Woodruff, J. L.	12	87	+ 7	5.08				
Dickerson, J.	10	350	— 1	4.99				
MacFarland, J. George, R. W.	1	470	+ 98	5.20	929	— 185*	4.80	Double witnessed and rechecked. Work done at Tarkio College.
Total*	92	1774	+ 319*	5.17	954	— 175*	4.81	

C. Sealed Key Decks

Pratt, J. G. Price, M. M.	12	40	+ 24	5.60	.....	.....	.....	Double witnessed and rechecked. Work done at Oxford Orphanage.
Weckesser, M.	11	232	+ 41	5.17	.....	.....	.....	Double witnessed and rechecked.
Gibson, E. P.	1	10	— 5	4.5				
Total*	24	282	+ 60	5.21				

*For grand totals of these sections of Table I see Table II.

TABLE II

TOTALS FOR ALL UNRESTRICTED SHUFFLE WORK

Classification	Experimenters	Subjects	Runs	Dev.	Av/25	S. D.	C. R.
Unscreened	9	129	5451	1405	5.25	150.6	9.3
Screened	9	92	2728	494	5.18	106.5	4.63
Sealed	4	24	282	60	5.21	34.25	1.75
Total*	13	203	8461	1959	5.23	187.7	10.40
Total Screened and Sealed	11	104	3010	554	5.18	112	4.94
Total Double Witnessed and Screened or Sealed	10	50	2061	424	5.20	92.6	4.57
Total Rechecked and Screened or Sealed	3	24	1671	348	5.20	83.4	4.17

*See text, page 121.

DISCUSSION

The "ESP Shuffle." The size of the critical ratios obtained from the results reviewed above designates the test performances as clearly extra-chance in character. The conditions, particularly those in which the target pack was screened throughout the test from sensory contact by the subject, indicate quite definitely the extra-sensory character of the perception upon which the performance was based. The number of the experimenters, along with the practice of double-witnessing and of double-checking in the recording and handling of large and important sections of data, give strong assurance that persistent error is not the explanation for the results. It would appear that extra-sensory perception has played a part in the shuffling of the cards by some of these subjects under the conditions described.

The favoring of alternatives to the hypothesis of the ESP shuffle would of course greatly strengthen the evidence for the precognition hypothesis submitted in the preceding report, "Pre-Shuffle Card Calling." That is to say, if the ESP shuffle hypothesis is supported, it weakens the pre-shuffle calling evidence of precognition. If it is a poor hypothesis, then automatically there is no good alternative to precognition to account for the results published. It will be recognized by all that the ESP shuffle is a simpler, more parsimonious hypothesis than that of precognition.

On the other hand, as pointed out above, alternatives to the ESP shuffle hypothesis are not easy to find. Chance, sensory cues, and cumulative error are all especially recognized and met. All the statistical requirements are duly regarded, as for example, the avoidance of

selection of results by the inclusion of all data obtained under the test conditions, and the separate evaluation of subdivisions only where there was a clear-cut difference in the conditions.

But certain ill-defined hypothetical alternatives still hover on the edge of formulated expression: Is it possible that card peculiarities, similar from pack to pack, might lead to similar points of easier cutting and thus cause occasional extra-chance coincidences? Could personal similarities in subject and experimenter, in habits of shuffling, cutting, or otherwise handling the cards, enter into the effect obtained? Or could any special grouping of certain cards in the pack, possibly resulting from similarity in shape due to warping (occurring in the cards from the point of manufacture, from rough handling in the tests, or from deliberate trickery on the part of the subject) produce positive extra-chance deviations?

These, and perhaps other vaguely defined possibilities, are excluded from explanation of present results by certain features of the evidence. For one thing, these alternatives all depend upon two-pack similarities; but in the George and MacFarland series, the matching was made against screened symbol lists (calls made earlier by other subjects) instead of against another pack.

Also, the experimenter's pack in much of the best work, particularly the two-experimenter series and the sealed pack series, could not have been tampered with by the subjects (many of whom were children, and among these a considerable portion were wholly or in part blind).

Finally, the general procedure of pack-matching has been done so extensively in control experiments on probability theory that, were there to be some common extra-chance factor to be expected other than that under test (ESP) it would in all reason be expected to have appeared in these control series. The chance control matchings were conducted much as in the tests here reported, with the all-important exception that the subjects (and the experimenters) had no knowledge of, nor interest in, the ESP shuffle hypothesis. That is, they were not disposed or motivated to exert any but a random effect on the card shuffling and were not aware that ESP might play any determinative role. One other difference existed; namely, that in the control series there was as a rule a fixed minimum number of shuffles. This difference, however, will not invalidate the comparison as bearing on the type of alternative hypotheses under consideration here (especially so since the experimenters in the work here reported likewise used a certain minimum number of shuffles, though of course the subject did not).

The most extensive pack-matching chance control series is that of Warner, reported in his article, "The Role of Luck in ESP Data,"[5] involving the matching of 4,383 packs. The result was well below the requirement for a significant deviation; the average per 25 was 4.95. There have been other shorter series, published and unpublished, with results well within the range of chance.

The hypothesis, then, briefly represented as an "ESP shuffle" and which arose out of the pre-shuffle card-calling method of testing for precognition, has apparently a basis in empirical fact. It is therefore not unreasonable to hold that an experimenter might conceivably shuffle a pack of cards in a precognition experiment, expecting to give the pack a thoroughly random rearrangement and yet at the same time unwittingly give the cards in the pack something other than a random order; in other words, by means of ESP to arrange the pack of cards to some slight degree so that it will conform in a more than accidental way to the series of calls which were intended as predictions.

Restricted Shuffling Series. However, a further important step remains; namely, to determine whether under the more exact conditions (i.e., limited shuffling) used by the experimenter in the precognition series the shuffling can be done to a degree that is significantly beyond chance expectation and is equal to the scoring obtained in the precognition series.

Experiments with limited shuffling have already been started by a number of investigators, and a total of 1,810 runs have been reported to date. These, however, have yielded a total positive deviation of only 89; and it is desirable that further work be done before any conclusion is reached. The results at this stage would appear to indicate that the conditions of the restricted shuffling which were used in the precognition tests do not allow the functioning of ESP in guiding the shuffling to a degree that would be reliably above chance. If we examine the sub-series of this limited shuffle work, however, we observe that one set of experimenters, George and MacFarland, have obtained a significant positive deviation on the 115 runs which they have reported to date. This series gives a critical ratio of 3.2. Furthermore, the total of the screened series of 1,300 gives a positive deviation of 128 with a critical ratio of 1.7 (see Table III). These facts indicate that it would not be wise to draw a conclusion at the present time. These preliminary and incomplete results are mentioned to indicate the direction of further research and to suggest restraint in drawing conclusions on this point.

[5] J. PARAPSYCHOL., 1937, 1, 84-92.

TABLE III

RESULTS OF TESTS WITH RESTRICTED SHUFFLE

Experimenter	Subjects	Time Limit	1 Cut	1 Shuffle	3 Shuffles	4 Shuffles	5 Shuffles	6 Shuffles	Conditions
Gibson, E. P.	1	42+2	.....	.....	.....	.....	.....	58—1	Unscreened
Dickerson, J.	10	.....	.....	330+39	300+21	.....	295+21	.....	Screened**
Smith, B. M.	4	64+4	64—13	64—16	64+6	.....	64—6	.....	Screened**
MacFarland, J. George, R. W.	4	.....	.....	.....	.....	115+72* (Double Witnessed)			
MacFarland, J.	.....	.....	.....	.....	.....	115—12			
Weckesser, M.	.....	100+9	.....	.....	.....	135—37	.....	.....	Sealed
Totals	.....	206+15	64—13	394+23	364+27	365+23	359+15	58—1	

Grand Total.....1,810 Runs +89
S. D. = 86.78
C. R. = 1.02

*Significant, C. R. = 3.2
**Total screened is 1,360 runs. C. R. = 1.7

In other words, it would not appear fully warranted as yet to conclude that restricting the shuffling does or does not prevent the operation of ESP in influencing the order of the cards in the pack and in the absence of this conclusion the ESP shuffle hypothesis still remains logically a possible alternative to the precognition hypothesis insofar as the research reported to date is concerned.

However, in the conclusion of the precognition report it was pointed out that, assuming the ESP shuffle hypothesis to be applicable, there were certain instances in which it could not account completely for the results of precognition tests; that is, the case for precognition need not be regarded as conclusively swept away. It must, however, be conceded that the case is greatly weakened by the tests of the ESP shuffle hypothesis reported in this paper and that it would be hazardous to credit the precognition hypothesis with confidence until evidence of other character is forthcoming.

The next report in this series will attempt to take a further step toward a crucial testing of the precognition hypothesis by presentation of evidence designed to evade the alternative of the ESP shuffle hypothesis.

SUMMARY

As a result of tests of the precognition hypothesis in which a preshuffle card-calling method was used, the alternative arose that perhaps the experimenter was able to shuffle the pack in favor of the subject's calls by means of extra-sensory perception of the card order. Investigation of this possibility was made by asking the subject deliberately to shuffle one pack of cards to match another, both packs being inverted

and with faces of the cards invisible. In 5,451 such pack-matchings, and in 3,010 pack-matchings in which the key pack (or series of symbols to be matched) was screened or sealed from view, the deviations from the mean in the intended direction were so great as to rule conclusively out of consideration the hypothesis of chance coincidence. The conditions, particularly with the screened or sealed key pack, excluded sensory perception; and double-witnessing and double-checking give additional assurance of accuracy for large blocks of the data presented. The indications are that the hypothesis of a role for ESP in the shuffling of a pack of cards is experimentally demonstrated.

An experimental series to determine whether the ESP shuffle can operate significantly under conditions of a restricted shuffling is inconclusive, but the results are sufficient to suggest that the ESP shuffle may be found effective even under those conditions. This experimental series continues. Although the previous precognition report is considerably qualified by the ESP shuffle work, another approach to the precognition hypothesis will be undertaken in the third article of this series.

ON THE COMBINATION OF CRITICAL RATIOS

FRANK W. BUBB

ABSTRACT: In any extended ESP investigation the data naturally accumulate in a series of subsets, one for each experiment. For each of these parts a critical or deviation ratio is usually calculated. It becomes convenient then in viewing the investigation as a whole to have a formula for combining these sub-ratios into a single composite ratio. This paper sets up such a formula, upon a strict statistical basis, for combining deviation ratios. This formula is compared critically with Dr. Rhine's empirical formula and the latter is shown to give, for the data to which he applied it, a close approximation to exact values. A remarkable connection is found to exist between Rhine's empirical formula and the theoretical formula.

Dr. Bubb is head of the Department of Applied Mathematics in Washington University, Saint Louis.

THE INVESTIGATION of ESP has been guided by Professors McDougall and Rhine along lines which permit of mathematical treatment. Statistics provides a choice of several more or less appropriate measures for expressing the bearing of any of these experiments upon the existence of ESP. Dr. Rhine selected for this purpose the deviation ratio X, also called the critical ratio; and this choice has been so generally accepted that the correct evaluation of this ratio X is one of the first aims of any experiment which follows the Duke tradition.

In any extended investigation, the data naturally accumulate in a series of subsets, one subset for each partial experiment. For each of these parts a critical ratio is usually calculated. It appears to be convenient then in viewing the investigation as a whole to have a formula for combining these partial ratios into a single composite ratio. It is the purpose of this paper to set up such a formula for combining critical ratios. The formula resulting from this argument is then compared with the formula

$$X = \sqrt{X_1^2 + X_2^2 + \ldots\ldots + X_s^2} \qquad (R)$$

which Rhine proposed[1] upon a frankly empirical basis and which he and others have used in evaluating their data.

The argument will be based, of course, upon the chance hypothesis. The notation follows:

$n =$ number of individual card trials.
$m =$ number of correct calls or hits.
$p =$ probability of a correct call.
$q =$ probability of an incorrect call.
$pn =$ most probable number of hits.
$D =$ deviation $= m - pn$.
$\sigma =$ standard deviation $= \sqrt{pqn}$.
$E =$ probable error $= 0.6745\ \sqrt{pqn}$.
$X =$ critical or deviation ratio $= D/E$.

The last definition is equivalent to the formula

$$X = (m - pn)/0.6745\ \sqrt{pqn}. \tag{1}$$

Consider a series of s experiments. We assume that p and q remain constant throughout all the s experiments; for example, at the values 1/5 and 4/5 as in the standard ESP card experiments. For the α-th experiment then, we have by the definition (1) the s equations

$$X_\alpha = (m_\alpha - pn_\alpha)/0.6745\sqrt{pqn_\alpha} \quad (\alpha = 1,2,\ldots,s) \tag{2}$$

The following conditions should be noted at this, their point of origin:

$$\left.\begin{array}{l} \text{If } m_\alpha > pn_\alpha \text{ then } X_\alpha \text{ is positive.} \\ \text{If } m_\alpha < pn_\alpha \text{ then } X_\alpha \text{ is negative.} \end{array}\right\} \tag{3}$$

The composite X, that is, the critical ratio calculated by pooling the s subsets to form a single composite set of data, is also given by definition (1). Thus,

$$\left.\begin{array}{l} X = (m - pn)/0.6745\ \sqrt{pqn} \\ m = m_1 + m_2 + \ldots + m_s = \Sigma m_\alpha \\ n = n_1 + n_2 + \ldots + n_s = \Sigma n_\alpha \end{array}\right\} \tag{4}$$

From these equations we readily obtain by regrouping the terms,

$$X = \Sigma(m_\alpha - pn_\alpha)/0.6745\ \sqrt{pqn}. \tag{5}$$

By inserting into (5) the values of $m_\alpha - pn_\alpha$ found by solving (2), we find, upon cancelling the quantities $0.6745\ \sqrt{pq}$, the formula

$$\begin{aligned} X &= (X_1\sqrt{n_1} + X_2\sqrt{n_2} + \ldots + X_s\sqrt{n_s})/\sqrt{n_1 + n_2 + \ldots + n_s} \\ &= \Sigma X_\alpha \sqrt{n_\alpha}/\sqrt{n} \end{aligned} \tag{6}$$

[1] See *Extra-Sensory Perception* by J. B. Rhine, p. 34. Also "Mathematical Techniques Used in ESP Research" by J. A. Greenwood and C. E. Stuart, J. PARAPSYCHOL., 1937, vol. I, p. 206.

In order to simplify further discussion, let us define the quantities

$$W_a = \sqrt{n_a/n} \quad (a = 1, 2, \ldots, s) \tag{7}$$

as the critical weight factors for the respective experiments. The set of W's form a normalized set, namely, they satisfy the equation

$$\Sigma W_a^2 = W_1^2 + W_2^2 + \ldots + W_s^2 = 1 \tag{8}$$

and furthermore, each one of the W's must, in accord with (3) be taken as positive. Using these critical weight factors, the equation (6) becomes

$$X = \Sigma W_a X_a = W_1 X_1 + W_2 X_2 + \ldots + W_s X_s. \tag{B}$$

This is the required formula for calculating a composite critical ratio.

Let us apply the correct formula (B) and the empirical formula (R) to an actual example—let us say to the data given in TABLE XXVII on page 89 of Rhine's book "Extra-Sensory Perception." The results are collected in the accompanying table. The columns marked trials and hits were taken from Rhine's table. All other data in the table were calculated.

Exp	Trials	Hits	pn	D	E	X	W	WX	X^2
1	10275	3937	2055	1882	27.35	68.82	.6500	44.72	4736
2	1750	596	350	246	11.29	21.80	.2682	5.85	475
3	6175	2220	1235	985	21.20	46.46	.5038	23.41	2158
4	1725	406	345	61	11.21	5.44	.2663	1.45	30
5	4400	1340	880	460	17.90	25.70	.4253	10.93	661
	24325	8499	4865	3634	42.08	86.36		86.36	8060

By Formula (R), $X = \sqrt{8060} = 89.78$

This table exhibits three ways of calculating X. Thus, applying the definition (1) directly to the pooled data in the last line of the table, we have

$$X = (8499 - 4865)/0.6745\sqrt{.2 \times .8 \times 24325} = 86.36$$

and this number is tabulated at the foot of the column X. The formula (B) is applied simply by adding the figures in the column marked WX, and again the value for X is 86.36, which checks, as it should, the former value. And finally, in applying the formula (R), the last column marked X^2 is summed and then the square root is taken; this gives the value 89.78. It appears then that formula (R) overestimates the value of the composite X by about 4% in this case. While it is true that this does not alter the astronomical character of Rhine's proof for the existence of ESP, the error is nevertheless in the wrong direction for the most conservative evaluation of results.

A further comparison of formulae (B) and (R) is brought out by the graphs in Fig. 1 which exhibits the results for the case of two composed experiments. In this case $X = W_1X_1 + W_2X_2$ with the constraint $W_1^2 + W_2^2 = 1$. Eliminating W_2 and making the change in notation: $X_2/X_1 = k$, $X/X_1 = y$, $W_1 = x$, we have

$$y = x + k\sqrt{1 - x^2}. \tag{9}$$

Fig. 1 shows graphs of this function for several values of k. It will be observed that the composite critical value X (which is proportional to y) varies markedly with changes in the relative number of trials in the partial sets. In every case the formula (R) gives the topmost

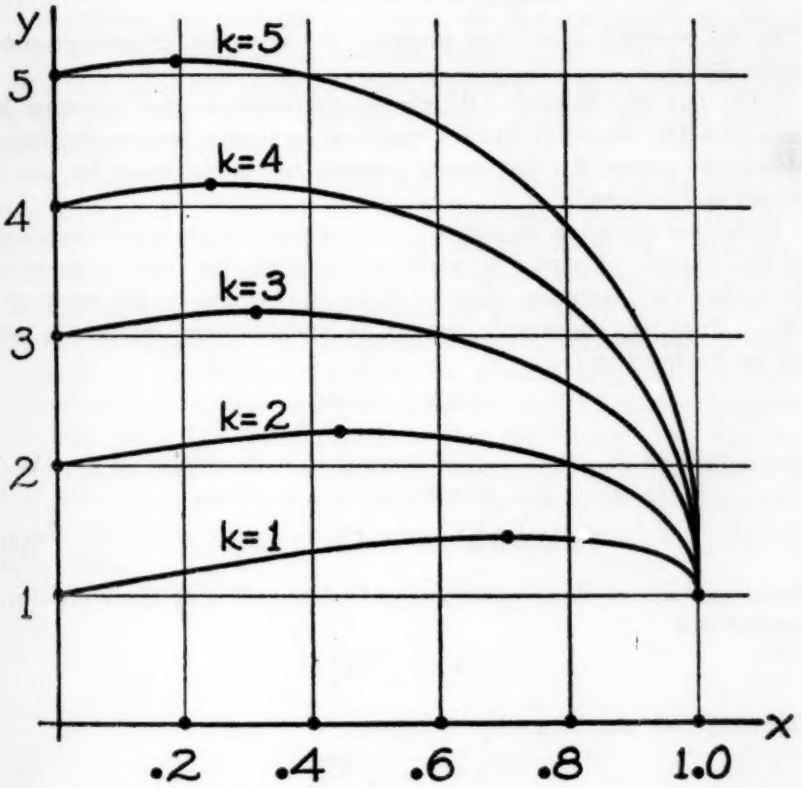


FIGURE 1.

point marked with a circle on the graph, in other words, the maximum value. This can be shown more generally.

In (9) the variable x is limited to numbers which are the square roots of positive rational numbers. In the case of large numbers, how-

ever, we can apply the methods of analysis. This may be justified by an argument which parallels the deduction of the Gaussian distribution as an asymptotic form of the binomial distribution. Such arguments are well known and the application here is obvious. Accordingly, we calculate the maximum value of (9) by equating its derivative to zero. This gives $x_{max} = 1/\sqrt{1 + k^2}$, $y_{max} = \sqrt{1 + k^2}$, or translating into the original notation,

$$\begin{aligned} W_1 max &= X_1/\sqrt{X_1^2 + X_2^2} \\ W_2 max &= X_2/\sqrt{X_1^2 + X_2^2} \\ X_{max} &= \sqrt{X_1^2 + X_2^2} \end{aligned} \tag{10}$$

The last equation shows that formula (R) gives the greatest possible value for X.

The fact that formula (R) always overestimates the composite X was probably hidden in Rhine's empirical treatment because the function is, of course, flat and nearly constant within the range for which he set up his formula.

The like result can be shown for a composite experiment consisting of any number of parts. In mathematical terms the problem amounts to finding the maximum value of X in (B) subject to the constraint (8). Applying Lagrange's method of undetermined multipliers, we set up the function

$$\Sigma W_\alpha X_\alpha - \lambda \Sigma W_\alpha^2$$

and equate all its s first partial derivatives with respect to W_1, W_2, ... , W_S to zero. This gives the set of s equations,

$$X_\alpha = 2\lambda W_\alpha \quad (\alpha = 1,2,\ldots,s) \tag{11}$$

Squaring each of these equations and adding all s of them together, we find that

$$\lambda = \sqrt{\Sigma X_\alpha^2}/2$$

Inserting this value of λ into (11) gives

$$W_\alpha = X_\alpha/\sqrt{\Sigma X_\alpha^2} \tag{12}$$

which are the values of W which make X a maximum. Since all these values are to be taken as positive, perhaps we ought to write (12) more explicitly in the form

$$W_\alpha = |X_\alpha/\sqrt{\Sigma X_\alpha^2}| \quad (\alpha = 1,2,\ldots,s) \tag{13}$$

The actual maximum of X is found by inserting these s values (13) into (B). The result is

$$X_{max} = \Sigma X_a |X_a| / \sqrt{\Sigma X_a^2} \qquad (14)$$

In case all the X_a are positive, this reduces to the form

$$X_{max} = \sqrt{\Sigma X_a^2} \qquad (R)$$

which is exactly Rhine's empirical formula (R). This is rather remarkable, because one rarely finds so close a connection between theory and an empirical formula.

No attempt will be made to complete the above argument by examining sufficient conditions for extreme values of X. Such a discussion would be lengthy and appears to meet no present need in ESP research. In any case, such extreme values have no bearing upon the general applicability of formula (B).

In closing, it should be pointed out that, although Rhine's empirical formula (R) overestimates the composite ratio—giving in fact, the optimum value in all cases—this does not invalidate his conclusions. In the first place, a recalculation of his results shows that in most cases his data locates a point near the maximum value of X where the function (B) is necessarily flat. In the second place, the astronomical magnitude of his probabilities would not be affected materially by much greater errors than those implied by the differences between formulae (B) and (R). Dr. Rhine points out that his formula (R) has two virtues which led him to adopt it: it is reasonably accurate and it is easy to manipulate. May I point out that the simple bilinear form (B) is still more easy to manipulate than (R). It also has the virtue of being exact.

ANALYSIS OF A LARGE CHANCE CONTROL SERIES OF *ESP* DATA

J. A. GREENWOOD

ABSTRACT: To check experimentally a number of the statistical questions arising in ESP research a random series of 500,000 call-card matchings was set up in such a way as to approximate a free-calling ESP test. The only point of difference was the careful exclusion of all possible extra-sensory factors in the results. Matching, binomial, normal and Pearson Type I hypotheses of distribution are applied to the resulting hit-frequency-per-run distribution. The distribution is that of a Type I curve. The best theoretical approximation is that of the binomial hypothesis.

Dr. Greenwood, of the Department of Mathematics, Duke University, will use this unusually large empirical series as the basis for further mathematical studies of ESP experiments.

THE REALIZATION in practice must in the last anaylsis determine the validity and usefulness of a fundamental hypothesis in probability or statistics. Especially is this so when the hypothesis in question is admittedly an approximation to the actual situation.

If one shuffled deck of ESP cards (25 cards, 5 of each symbol) were matched against another ESP deck a great number of times it would be no occasion for surprise to obtain a quite good fit to the theoretical matching frequencies of hits per run (4). Likewise, if the conditions patently approximated the binomial hypothesis in all essentials one would expect and probably get (chance alone obtaining) a quite good fit to the binomial histogram.

In the free calling experiments, which predominate the research so far, neither of the above conditions obtains. Nor is the theoretical distribution of hits for this case known. And it may be further observed that there would undoubtedly be a new distribution for each new call-frequency. That is, one would expect relative calls of symbols in number 8-8-7-2-0 to produce a different theoretical hit frequency distribution than, say, 66553. However, they all produce the same theoretical average number of hits. The exact variance for each case is known, also, and is less than that for the 55555 case (2).

Knowing, then, the variance of certain call series to be at most

but slightly in excess of that of the binomial, and that the matching frequencies are for all practical purposes identical with the binomial (4 and 6) we can only say that, not knowing contrariwise, our expectation is high that a nearly binomial distribution of hits would be produced.

It was with this in mind that the writer undertook the formation of just such a series with all known factors but chance ruled out as far as possible. Considering the extent and probable importance of the series no possible safeguard against error of any sort was omitted. As the usefulness of such a series is a direct function of its accuracy a detailed account of its derivation will now be given.

METHOD

It was decided to use the first twenty recorded call runs each of five of the best subjects at Duke as a sample call series. A copy of these was accordingly produced. It was only after the check series was well under way that the variance of this 100 runs was found to be 402.9833, an average per run of 4.029833, corresponding to a standard deviation (S.D.) of 2.00744. The binomial S.D. is of course 2.*

The 100 call runs were copied in columns, ten to the page on the convenient ESP score pads. Thirty decks of ESP cards (used throughout the work) were then shuffled in the following fashion. First, their order was reversed by the author. They were then passed to a second person who thoroughly mixed the cards in each deck, and then to a third person (ignorant of the use to which the cards were to be put) who gave each deck three dove-tail shuffles and a cut. This ritual seemed the best method at hand to avoid various ESP hypotheses.

With the help of a competent witness the card orders were copied as they were needed, one deck at a time, in a column of an ESP score pad. When the card order was copied the two experimenters counted the recorded symbols to make certain there were five of each. The column of card symbols was then cut out in a thin strip which could be placed next to the various recorded call series one after the other through the hundred. The two observers then counted the hits per run (the writer running down the column with a pencil) and recorded the total of such per run on a separate sheet. The used deck was placed separate from the unused ones. As the thirty deck boxes were numbered from 1 to 30 and used in that order, there was no possibility of an unshuffled deck being used a second time.

After 4,000 runs the ESP symbols were replaced by code numbers

* For the method used see ref. 2. Also note Table I.

12345 as the latter were found to be much easier to check. At the end of 7,000 runs a random sampling recheck of 200 runs netted two errors of omission, lowering the score by one in each case. It was therefore decided to recheck the whole 7,000 runs or 175,000 separate comparisons. Accordingly, the writer went down one group of results and one or two assistants from the Parapsychology laboratory checked on another part. The procedure was to obtain the new result first and then compare it with the old record. All told, 81 mistakes were discovered, 72 of omission by one and 9 of additions to the score by one.

From runs 7,001 to 20,000, the end of the series, the same general procedure was used with the exception that the assistant and writer made *independent* records of scores, working on different parts of the series. At intervals the two records were then compared for differences. If differences were encountered these particular runs were carefully rechecked and errors corrected.

THE QUESTION OF ERRORS

While there were three persons and sometimes four who observed every entry of the first 7,000 runs, there is every reason to believe that with only two observers under the new system the last 13,000 runs are just as reliable.

In the checkup referred to above the method was as follows. The assistant (at a distance precluding seeing the second record) read aloud his scores, the writer silently comparing them with his own results. As soon as a dissimilarity was observed the assistant was notified of the fact but not of the nature of it. Since it was noted that the assistant made more errors than the writer, it was left to the assistant to re-obtain his scores at those points. If the disagreement persisted the writer then re-obtained his own scores before changing the record. In this group the mistakes found in the writer's record were but five of omission of one hit, three of addition of one hit, and one of addition of two hits; nine errors in 13,000 runs or 375,000 calls.

The reader may raise the question in his own mind as to the likelihood of equal errors occurring simultaneously in the two records, thus escaping detection. A reasonable assumption to make is that the number of detected errors exceeds the undetected ones. However, let us momentarily deny this by assuming that there were, say, 20 errors, detected and undetected in one record, and 100 errors in the other record. The probability that each record would have an error in a preselected run is 20/13,000 and 100/13,000, respectively,—that they should occur together at the same place is $2{,}000/(13{,}000)^2$. The expectation of co-

incidental errors is then 2,000/13,000 or .15 errors. And this figure should be decreased when we realize that the errors are not all of the same direction and magnitude. Thus essentially *all* the errors were detectable ones. This shows clearly that with the detectable errors corrected as described, we are morally and practically certain of a large check series free from error.

For the whole series there were 12 errors of addition of one hit to a run, one of addition of two hits, and 77 of omission of one hit in the author's record. The total was 90 errors lowering the score by 63 hits out of an expected 100,000 hits. Since an urgent consideration of this series was the avoidance of errors on even the first going over, it would seem appropriate to stress the fact that with a little less rigor the errors would undoubtedly have been much more numerous. In particular, since omissions were the more common type of error, work which gives rise to a significant *negative* deviation should call for rechecking. This type of error would of course only serve to diminish a positive deviation and the misleading effect be at least a safer one.

ANALYSES

Symbol Preferences. The general effect upon distribution of scores due to subject symbol preferences must, according to probability theory, be a function of the relative numbers of the frequencies of symbols called, not of the actual symbols chosen, themselves. Consequently, Table I shows only the numbers of each symbol called per run of the 100 chosen call runs which formed the basis of the series. For example,

TABLE I

PATTERN TYPES WITHIN THE CALL SERIES

Type	Freq.	Type	Freq.	Type	Freq.
12,6421	1	87541	1	76660	1
11,6530	1	87532	1	76651	2
10,7521	1	87442	1	76633	3
10,6531	2	87433	2	76552	3
10,5532	1	86551	1	76543	11
10,5433	1	86542	3	76444	2
98332	1	86533	4	75553	3
96532	2	86443	1	75544	5
96442	1	85543	4	66652	2
95542	1	85444	3	66643	1
95533	1	77551	2	66553	9
95443	1	77542	1	66544	4
87631	1	77533	1	65554	8
87622	1	77443	4	55555	1
Total Freq.					100

the entry 86542 with frequency 3 means that 8 of one symbol, 6 of another, etc. through the run were called and that this pattern was true of three runs. Note that the 8, for example, might be associated with different symbols for different runs.

In this paper a study of the combined effect, only, of the 100 runs is made. At a future date it may be that the results due to certain selected runs of the call series will be singled out for further analysis.

Of the results of the check series the observed frequencies were transferred to 80 score sheets bearing 250 run scores each. Each sheet was tallied (twice, independently) to form 80 histograms of the observed frequencies. The 80 histograms were then summed to give the total hit-frequency distribution in the "observed" column of Table II.

This table displays not only the observed frequencies but also the corresponding expected frequencies under the respective hypotheses: binomial $(.8 + .2)^{25}$, normal $f(x/S.D.) = f(x/2.00744)$, matching, and Type I curve. Under each column is given the probability obtained by a chi-square Goodness of Fit test of the observed data to the corresponding hypothetical distribution.*

In Table III is given, as an example, the Goodness of Fit test of the data to the binomial distribution.

TABLE II

COMPARISON OF OBSERVED DATA WITH VARIOUS HYPOTHESES

Hits	Obs.	Expected			
		Binomial	Normal	Matching	Type I Curve
0	82	75.56	235.4	85.71	73.31
1	482	472.24	546.0	509.04	476.23
2	1,426	1,416.71	1,301.2	1,467.27	1,463.73
3	2,818	2,715.36	2,419.6	2,731.43	2,787.26
4	3,761	3,733.62	3,510.8	3,686.47	3,761.97
5	3,871	3,920.30	3,974.6	3,838.86	3,882.57
6	3,240	3,266.92	3,510.8	3,205.15	3,205.10
7	2,139	2,216.84	2,419.6	2,201.32	2,176.11
8	1,238	1,246.97	1,301.2	1,266.21	1,234.38
9	588	588.85	546.0	617.93	589.68
10	253	235.54	178.8	258.26	237.70
11	71	80.30	45.8	93.07	80.55
12	25	23.42	9.2	29.06	22.74
13	5				
14	1	7.38	1.6	10.21	6.41
15	0				
Total	20,000	20,000.01	20,000.6	19,999.99	19,997.74
		P = .58801	$\chi^2 = 180.17$ $n' = 8$ $P = 10^{-35}$	P = .11845	P = .8835

* For a discussion of the chi-square tests applied to ESP data see ref. 3.

TABLE III

χ^2 FIT OF OBSERVED SERIES TO BINOMIAL

Hits	Expected Frequency	Observed Frequency	$(f - f_0)^2/f$
0	75.56	82	.55
1	472.24	482	.20
2	1,416.71	1,426	.06
3	2,715.36	2,818	3.88
4	3,733.62	3,761	.20
5	3,920.30	3,871	.62
6	3,266.92	3,240	.22
7	2,216.84	2,139	2.73
8	1,246.97	1,238	.06
9	588.85	588	.00
10	235.54	253	1.29
11	80.30	71	1.08
12	23.42	25	.11
13-25	7.38	6	.26
Total	20,000.01	20,000	

$\chi^2 = 11.26$
Degrees of freedom $= 13$
$p = .588$

From the probabilities in Table III it is seen that the binomial is an excellent fit; that in a thousand repetitions of similar check series conducted under binomial conditions 588 would give a worse "fit" to binomial than the *actual* one. For application of a chi-square Goodness of Fit test to ESP data of this magnitude the normal hypothesis is quite unsatisfactory. The matching hypothesis can be said to be quite acceptable since 118 out of a thousand such series would give a greater chi-square than observed. The Type I curve is indicated and gives the best fit since its constants involve the empirical moments. Its equation (1, p. 54 ff.) is

$$y = 3{,}945.87\left(1 + \frac{x}{7.3351}\right)^{9.32424}\left(1 - \frac{x}{20.3362}\right)^{25.85097}$$

A comparison of statistics of the various distributions including those of the observed series will be found in Table IV.

It will be observed that with the exception of the mean all the essential statistics of the observed data are quite coincident with those of the binomial, and all, including the mean, are well within any range of deviation accepted as chance. Since the empirical distribution fits the binomial curve quite closely according to the chi-square Goodness of Fit test, the results conform also to hypotheses of independence of trials and constant probability 1/5 for each trial. This is further con-

TABLE IV

COMPARISON OF STATISTICAL CONSTANTS

Statistic Name	Value				σ of * Statistic	Critical Ratio of Difference between Observed and Binomial Statistic	Probability
	Observed	Binomial	Matching	Normal			
Average.....	4.9743	5.	5.	5	.01419	1.811	.0701
Standard Deviation....	2.0058	2.	2.0412	2	.010	0.159	.8737
Asymmetry (Skewness)...	0.3131	0.3000	0.3195	0	.0173	0.7563	.4497
Flatness (Kurtosis)....	.00597	0.0100	0.0261	0	.0346	0.1163	.9077
Third Moment.....	2.5267	2.4	2.7174				
Fourth Moment.....	48.4675	48.16	52.5362				

*Under normal assumption. See T. C. Fry, *Probability and its Engineering Uses*, p. 470, for the formulae giving the entries of this table.

firmed by the Lexian ratio, empirical S.D. divided by binomial S.D. This ratio, $2.00585/2 = 1.00298 \pm .0050$, is insignificantly variant from 1, its value under a binomial hypothesis.

Table IV, in conjunction with the size of the series, emphasizes the fact that the hypothetical population, from which the check series is a sample, can with confidence be assumed to be binomial in character as far as any of the customary statistical tests are concerned.

The graphs of the five distributions that have been compared are displayed in Table V. Since the graphs are but images of the frequency distributions of Table IV no discussion of them need be made. The practical coincidence of four of the distributions is rather vividly shown there, however.

It will now be possible to answer in a substantial empirical fashion many questions that have been raised regarding the ESP tests. For example, the suggestion has been given that perhaps those subjects scoring far above results generally accepted as chance made a relative number of calls of each symbol per run peculiarly conducive to high scoring. This argument has been answered theoretically by proofs that the expected mean is always 5 and the variance never greater than 4.17, regardless of the call frequencies (3 and 2). That is, the theoretical chance dispersion of scores about a mean of 5 is limited by that of the matching of one ESP deck against another. In other words, a subject who gives frequency preference to certain symbols induces a theoretical frequency distribution for his results which is more

closely grouped about the mean, with a consequent loss of extreme scores.

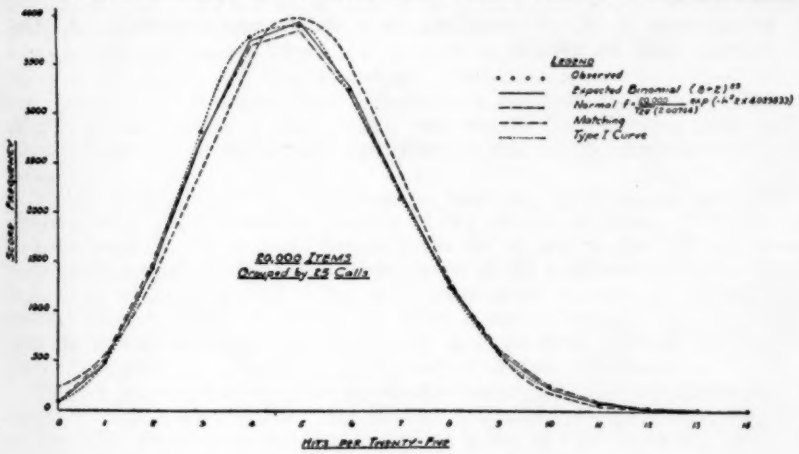
The 100 chosen call series produced the high scoring quite well as the following distribution of hits actually made by the five subjects with these call series shows.

Hits	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Freq.	0	0	1	4	2	6	15	10	17	14	10	10	5	1	3	1	0	0	0	1

With a theoretical S.D. of 20.0744 and critical ratio of 16.49, the chance probability is at least as small as 10^{-30}. Since over a period of 20,000 runs made under "pure chance" conditions this call series gave just chance results it can be quite safely concluded that there was nothing intrinsically conducive to extra-chance scoring in the relative frequencies of calls. This agrees with similar observations by Willoughby (7) and Martin (5).

TABLE V

GRAPHICAL COMPARISON OF OBSERVED SCORES WITH VARIOUS APPROXIMATIONS



CONCLUSIONS

From this construction of a large empirical chance series of scores comparable to those of ESP tests, the major conclusion is that the expected distribution of scores of free-calling experiments may be treated as binomial in character.

Previous observations that call series with which subjects obtained phenomenally high total scores are not necessarily intrinsically conducive to extra-chance scoring is herein substantiated.

Errors of checking scores noted in this series were predominantly

those of omission of correct correspondences, the tendency of which was to lower the true score.

(All the material of the data of this series is on file in the Parapsychology Laboratory of Duke University. It is of such a nature and so ordered that any of the results can be rechecked by anyone at any time without difficulty.)

REFERENCES

1. DAVENPORT, C. B. AND EKAS, M. P. Statistical methods in biology, medicine, and psychology. New York: John Wiley & Sons, Inc., 1936.
2. GREENWOOD, J. A. Variance of the ESP call series. *J. Parapsychol.*, 1938, **2**, 60-64.
3. GREENWOOD, J. A. AND STUART, C. E. Mathematical techniques used in ESP research. *J. Parapsychol.*, 1937, **1**, 206-225.
4. GREVILLE, T. N. E. Exact probabilities for the matching hypothesis. *J. Parapsychol.*, 1938, **2**, 55-59.
5. MARTIN, D. R. Chance and extra-chance results in card-matching. *J. Parapsychol.*, 1937, **1**, 185-190.
6. STERNE, T. E. The solution of a problem in probability. *Science*, 1937, **86**, 500-501.
7. WILLOUGHBY, R. R. Prerequisites for a clairvoyance hypothesis. *J. Appl. Psychol.*, 1935, **19**, 543-550.

LETTERS AND NOTES

Comments by Dr. Kellogg:

To the Editors:

I have been looking through No. 4 of your JOURNAL, and have a few comments to make on the Stuart & Greenwood article, to clear up minor errors and misunderstandings.

1) The section on "Mean Chance Expectation" gets a little off the track, towards the close. It is quite correct that the linkage between events in the run, treating the calls as roughly equivalent to matching by a second pack, is not such as to affect the average expectation. The next to the last paragraph, however, is phrased so as to imply a denial of the linkage within the single run, as shown in section 2 of my "Matching and Sampling." The probability is not exactly 1/5 on each draw, within any one run.

2) In the section on "Stand. Deviation," the second of the two italicized sentences at the close of the first paragraph does not quite describe the case, at any rate for me, and I am linked with it by being quoted in the next paragraph. My point was not that the variance was so much greater as to make the critical ratios invalid, but merely that it was greater, how much so being unknown at the time of writing, and that, accordingly, the matter should be studied before any conclusions were drawn. Now that the study has been made, it is obvious to all that the discrepancy is slight. The treatment of the conditions under which the difference should be allowed for, given in this article, seems to cover the ground.

3) The brickbat including the quotation from me, at the top of page 301, is accepted as probably deserved, because of my clumsy phrasing. I appear not to have made it clear enough, though I did try to say so, that this increasing discrepancy applies to the more extreme parts of the complete theoretical distribution. It would be marked if the total scores to be evaluated represented high average success. Since deviations for large samples have all been very small, only the part of the range near the average is in question in actual practice, and there the discrepancy diminishes as the number of runs increases.

4) The later and more accurate calculations of the matching series will of course be used instead of mine, but perhaps it is still not certain just which set to use. Dr. Huntington sent me a copy of a list of the frequencies given by Brown, with relative frequencies derived by himself. Taking these as the standard, there are errors in the third decimal place in Sterne's table. So, if Sterne's expectation of convergence of Type I to the correct chances is valid, there must be *some* error in Brown's work. Practically, of course, any of these series represents the trend well enough.

5) Now for a really important point. The article represents me as holding out for the matching hypothesis, as if the discrepancy in the chances were of serious practical importance. The really important part of the problem is that it has been shown that each run is a compound event, and that, accordingly, when one wishes to evaluate the significance of the difference between the average of a set of experimental scores and the chance average, it is necessary to take into account the higher moments, at the very least the standard deviation, of the scores. Since the separate calls are not independent, the binomial theorem

is logically invalid, the formula for probable error of the difference is similarly invalid, and therefore critical ratios on that basis do not give any estimate of the meaning of the data. The tables setting forth the chances, including rating tables providing for rough estimation of the chances for total scores, have no bearing upon this problem of sizing up evidence as such. They simply indicate whether, under the conditions assumed, such and such an event would be frequent or rare. In the latter case, it is natural to seek some explanation other than chance. Accordingly, relationship to the chance series is no longer in question, but instead difference from it. We wish to know how much confidence in the obtained results is justified.

6) This last problem, together with that of the selection of evidence, constitute the statistical aspect of the research as a whole. Neither of these problems has been touched upon by any of the mathematicians—except the remarks on sampling on the last page of Stuart and Greenwood's article in No. 3. Accordingly, the wording of the manifesto by Professor Camp is seriously misleading, since it is the mathematical aspect of the research upon which substantial agreement has now been attained. It is quite clear that the problems of selection and treatment of evidence as such, and of rigorous control of the experimental procedures, are approaching solution, but are not yet fully settled.

Yours sincerely,

CHESTER E. KELLOGG.

McGill University
Montreal, P. Q.

* * *

In Reply to Dr. Kellogg:

To the Editors:

We are glad that Dr. Kellogg has taken this opportunity to clear up any misunderstandings of his views implied by our article. Our intention was to discuss certain areas of criticism rather than the whole thesis of any one critic. We tried to make this explicit, but are glad to restate that in no case does our discussion deal completely with the argument of any critic.

But we feel that Dr. Kellogg's letter only emphasizes the necessity of the point of view taken in our article, that mathematical statements about ESP research should be tested by mathematical proof, and should be amenable to empirical demonstration in an experimental situation comparable to the ESP tests.

In his paragraph 1, Dr. Kellogg points out that although the average probability is 1/5 "the probability is not exactly 1/5 on each draw, within any one run." Since we have *proved* that the probability (as used in the new edition of Yule, p. 336) is 1/5, regardless of the number of runs, and since empirical checks have verified this expectation, we feel that Dr. Kellogg's assertion will be found false whenever proof of it is attempted.

Since we have discussed the matter of "linkage" within the run in several articles, we do not mean to imply a denial of its existence. But we have shown that the intercorrelation does not change the mean chance expectation.

No brickbat was intended in our quotation from Dr. Kellogg's article nor do we feel that the difference of opinion is one of language. There *are* differences between the various hypotheses suggested for evaluating ESP data. Our task was to find the magnitude and the implications of these differences. The findings simply contradicted Dr. Kellogg's statements. We are sorry to have to point out that Dr. Kellogg's further statement, that our demonstration is true only for averages close to the mean, is again at fault. The demonstration holds for the total range of the hypothesis in question.

When our paper was written, Dr. Kellogg's matching series probabilities were available only in his *Scientific Monthly* article. The series was incomplete in

that figures for frequencies close to the mean were omitted. Among the figures presented were some in which the errors, although of no consequence as errors, were greater than the difference between the binomial and matching series which the figures were offered to demonstrate. Sterne's approximation was more accurate as compared with Greville's exact values (J. PARAPSYCHOL., March 1938). Dr. Bancroft Brown's figures are identical with those of Greville.

We cannot agree with Dr. Kellogg's denial of the use of binomial theory on logical grounds. The statistical differences between binomial and other approximations has been shown to be too small to be of experimental concern, and the table on page 142 of this JOURNAL verifies that observation.

The suggestion that higher moments than the mean should be presented for the observed test scores, in order that other differences than that of the mean may be studied, may well be tried out by some investigator. At the present time, however, it appears that variation of experimental conditions will be a more fruitful approach to the meaning of ESP data than refinement of mathematical analysis.

CHARLES E. STUART
J. A. GREENWOOD

Duke University

RESEARCH NOTES

Summary of Tests with Miscellaneous Subjects

The first requisite in the study of ESP is a subject who can demonstrate consistently and to a measurable degree sensitivity beyond that of the usual sense organs. In a search for such subjects we have made informal tests of anyone who showed sufficient interest to work for us. In most cases these tests were not conducted under the critical conditions employed in the more extended study of individual subjects later undertaken. In other words, the subject, experimenter and cards were all in one room. The hyper-aesthesia hypothesis may, therefore, be applicable. Nevertheless, in the matching tests fresh cards were frequently introduced, and in the GESP and DT work the cards were not in the subject's line of vision.

These data are being published not as having unusual merit or novelty, but rather to guard us against the charge that we have withheld some results if, later on, we publish more significant results using gifted subjects. Table I summarizes all card work done by us except when a subject ran through fewer than ten decks. (Treatment of these discarded cases—496 decks with a positive deviation of .5 S.D.—shows that their elimination has not affected our figures appreciably.)

TABLE I

SUMMARY OF ALL TESTS

Tests	Subjects	Runs	Hits	Dev.	S.D.	C.R.
GESP....	41	1497	7499	+ 14	77.38	+0.18
OM.....	40	2902	14791	+281	107.74	+2.6
BM.....	13	1172	6121	+261	68.61	+3.8
DT.....	36	2429	12230	+ 85	98.56	+0.8
TOTAL..		8000	40641	+641	178.89	+3.58

Higher scores were observed on matching tests than on GESP or DT. Only the former surpass the critical ratio of 2.5 S.D. When results for all four tests are combined the deviation is found to be 3.58 S.D. At one time or another,

several of our subjects approached or surpassed 2.5 S.D., but at the time this paper is being prepared the total scores for only two subjects exceed this criterion. One, because of his health, is not now available for further work, but it is expected that M. H. can be tested under more rigorous conditions. (Results of a test made subsequent to the preparation of this note are given in an article entitled "A Test Case." J. PARAPSYCHOL., 1937, Vol. I, pp. 234-238.)

Since many of our subjects were tested by two or more methods it has been possible to obtain a correlation of ability on the several tests. Because of the small number of cases the Rank Difference Method has been used and the value of r inferred (see Garrett, H. E., Statistics in Psychology and Education Longmans Green, New York 1926).

TABLE II

COMPARISON OF TESTS

Tests	Subjects	r
GESP and OM.....	20	.45
GESP and BM.....	10	.32
GESP and DT.....	13	.05
OM and BM.....	9	.84
OM and DT.....	13	.64
BM and DT.....	6	.31

The OM and BM tests seem most closely related. Table III indicates this more accurately than does the correlation coefficient. While correlations be-

TABLE III

COMPARATIVE SCORES ON OM AND BM

Subject	Variation in terms of SD	
	OM	BM
MH.....	+4.6	+14.1
MS_2.....	+1.3	+ 1.4
MC_2.....	+1.2	+ 1.2
EH_2.....	+ .6	+ 1.6
SH.....	+ .6	— .2
PR.....	+ .3	— .8
GW_1.....	+ .3	+ .6
MR.....	— .2	— 1.0
RH.....	— .3	— .3

tween such small groups are never reliable, we can hardly belittle the fact that in all six cases the coefficient is a positive one. Nor can we overlook the scores of M. H., which were significantly positive on all three forms of tests given her. It is more difficult to believe that these relationships are due to pure chance than to suppose that they exist because the several tests are measures of one ability.

LUCIEN WARNER.
MILDRED RAIBLE.

The Hypothesis of Deception

Certain friends of the research in extra-sensory perception have recently informed us of rumors and gossip in which broad implications were made that the subjects at Duke University and at other places were practicing deception in the ESP tests, "pulling the leg" of the experimenters, and that even when caught, these deceptions were deliberately withheld from the public by the investigators. Such rumors have also reached us by other routes, and one such has recently appeared in at least one Spiritualistic journal. They appear to have originated in psychic research circles having to do largely with professional mediums.

For the ESP experimenter in the psychological laboratory the problem of excluding deception on the part of the subject is of course an aspect of the broader problem of excluding the possibility of sensory cues. The goal of the experimenter is the complete exclusion of all possible sensory cues, which includes assuming the dishonesty of the subject. This goal led to the introduction and use of sealed opaque envelopes, opaque wooden screens, and the separation of the cards from the subject by distances and walls.

In other words, while our psychical research friends suspect *some* of the ESP subjects, it might be said the ESP investigators are on guard against them *all*, or, more exactly, anticipate the possibility of any one of the group of subjects tested being willfully disposed to deceive, and are prepared to make it impossible for any of them to do so.

To such an extreme has this attitude of safeguarding of the experiments against willful or unwitting deception been carried that no magician, whose profession it is to deceive is willing to attempt to work (as magician) under such conditions.

The answer, then, to all those who suspect the honesty of subjects is: If under the conditions deception is humanly possible, the conditions are not adequate to establish the degree of confidence required for so weighty a conclusion as the occurrence of extra-sensory perception. Consequently, if it be discovered (and we have no anticipation of it being so) that one or more of the subjects in the ESP research turns out later to be of questionable character, it will have no bearing upon conclusions based on conditions in which sensory perception was clearly excluded as a possibility. As a matter of fact, in two such experiments a professional magician acted as a subject; and yet there is no ground for discriminating against his results in the tests.

It is of course to be expected, with the hundreds of subjects dealt with, including wide ranges of groups, ages, and classes, that some indications of a will-to-deceive should occasionally appear, more particularly with children when rewards are offered. But it has long been the rule not to draw final conclusions of the occurrence of ESP, except on the basis of conditions which, even given a subject with intent to deceive, fully exclude all sensory cues. As has often been repeated, no research has been submitted* without conditions in which either the cards are screened from the subject, or distance and walls intervene in a sufficient portion of the series to safeguard the conclusions. Thus the ESP conclusions can be in no danger from this source.

Given any evidence or suggestion of an intent of the subject to deceive, and given a set of conditions allowing any possibility of its being successful, the results of the series by that subject and for those conditions would be excluded from all serious consideration as a matter of course: the mere possibility alone, however, is sufficient to bar data from acceptance as evidence of ESP.

A further point is frequently associated with this hypothesis of deception. It is vigorously urged in some quarters that when we encounter a case of willful

*Except one in which the experimenter was her own subject.

deception we should hold the spectacle up to public scrutiny and publish it quite as freely and frankly as we report the test results. The many contacts established by the Laboratory with a wide range of claims and claimants to extrasensory abilities has naturally led to our encountering a number of phenomena which on closer investigation proved to be fraudulently produced. We do not feel, however, that any good purpose could be served by the exposure, à la Houdini, of these instances. In many of them the subjects were the victims more of ignorance of the importance of scientific soundness, than of dishonesty.

In a word, a research project in ESP does not become of conclusive scientific importance until it reaches the point at which even the greatest will-to-deceive can have no effect under the conditions. This criterion is the very threshold of the research field. It leaves us under no obligation to concern ourselves either with the ethics of the subjects or with the morbid curiosity of a few individuals.

J. B. RHINE.

GLOSSARY

In order to avoid constant redefining of commonly recurring terms in papers appearing in this journal, the following definitions are submitted for convenient reference. In case of any discrepancy between Glossary and usage in the text of an article the latter should be followed. Words defined elsewhere in the Glossary are italicized in the text of the definitions.

AGENT: In tests for *telepathy*, the person whose mental states are to be apprehended by the *percipient*.

AVERAGE SCORE = AVERAGE: Average number of *hits* per *run*.

BT (BEFORE TOUCHING): The technique in which the top card of the face-down *deck* is *called* and, after being *called*, is laid aside for *checking* at the end of the *run*. Each card in the *deck* is treated in the same way.

BM (BLIND MATCHING): The technique in which the *subject matches* a *deck* of *ESP cards* to five *key cards* which are laid out face down before him in an unknown order. Unless otherwise stated, the order is also unknown to the experimenter.

CALL v.: To designate a card by naming one suit either orally or by indicating it manually (as pointing or writing).

CALL n.: The response described above; also, the resulting selection.

CHANCE: The complex of undefined causal factors irrelevant to the purpose at hand.

CHANCE EXPECTATION = MEAN CHANCE EXPECTATION: The most likely *score* if only *chance* obtains.

CHANCE AVERAGE: *Mean chance expectation* in terms of *average* per *run*.

CHECK: To determine a *score* after the completion of a *run* by comparing the order of the *subject's calls* with the order of cards in the *deck*.

CLAIRVOYANCE: *Extra-sensory perception* of objective events as distinguished from *telepathic* perception (of the mental or subjective events of another person).

CRITICAL RATIO = X VALUE: The observed *deviation* divided by the *standard deviation*.

DECK: Twenty-five *ESP cards*, five of each suit.

DEVIATION: The amount an observed number of *hits* or an *average score* varies from the *mean chance expectation* or *chance average*. A *deviation* may be total (for a series of *runs*) or average (per *run*).

DT (DOWN THROUGH): The technique in which the cards are *called* down through the *deck* before any are removed or *checked*.

EMPIRICAL CONTROL: An experiment which wholly or partially follows the main experiment with the exception that the conditions are designed to exclude the possibility of *ESP*.

ESP (EXTRA-SENSORY PERCEPTION): Response to an external event (perception) not presented to any known sense.

ESP CARDS: *Cards*, each bearing one of the following five symbols: Star, circle, three parallel wavy lines (called "waves"), rectangle, plus.

ESP SYMBOLS: See plate opposite page 1, this journal Vol. I, No. 1, March 1937.

ESP TESTS: A considerable number of techniques come under this heading, conveniently represented by initials, the principal ones being: *BT, DT, PT, GESP, BM, OM, STM*.

EXTRA-CHANCE: *Significantly* different from *chance expectation*.

GESP = GENERAL (OR UNDIFFERENTIATED) EXTRA-SENSORY PERCEPTION: A technique designed to test the occurrence of *extra-sensory perception*, permitting either *telepathy* or *clairvoyance* or both to operate.

KEY CARD: One of the five cards (where there are five suits) against which the cards of the test *deck* (i.e. *target cards*) in the *matching* tests (*OM, BM, TM,* etc.) are *matched.*

MATCHING: A form of *calling* in which a *target card* is placed opposite the *key card* which the *subject* selects to identify it.

OM (OPEN MATCHING): The technique in which the *subject matches* a *deck* of *ESP cards* to five *key cards* which are face up before him.

PERCIPIENT: The person who perceives. In this journal the common use will pertain to *ESP* rather than to sensory perception.

PT (PURE TELEPATHY) = TELEPATHY: The word "pure" emphasizes the exclusion of *clairvoyance.*

RELIABILITY: *Significance.*

RUN: The presentation and *calling* of 25 *ESP cards* or *symbols* in succession.

SCORE: The number of *hits* made in one *run.*

TOTAL SCORE: *Score* of any number of *runs.*

AVERAGE SCORE: *Total score* divided by number of *runs.*

SCREEN: An opaque barrier used between the subject and the card or agent. The main types of screens will be illustrated in this journal on their first introduction in print.

STM (SCREENED TOUCH MATCHING): The *touch matching* technique with an upright *screen* preventing vision by the *subject* of the handling of the cards by the experimenter.

SERIES: Several *runs* that are grouped consecutively or according to some other principle.

SIGNIFICANCE: The unlikelihood that a given event (or series) will not on the average occur by chance more often than once in 150 such events (or series). [Arbitrarily taken to mean a *deviation* in the expected direction such that the *critical ratio* is 2.5 times the *standard deviation* (or four times the probable error) or greater.]

STANDARD DEVIATION: A unit of measurement in statistical method; that *deviation* above and below *mean chance expectation* which is expected to include about ⅔ the *chance scores.* For *ESP cards,* S. D. = 2.04 $\sqrt{\text{no. of runs}}$.

SUBJECT: The person who is experimented upon. Most commonly the *percipient* in *ESP,* though also the *agent* in *telepathy.*

TARGET CARD: The card which the *percipient* is attempting to perceive (i.e. to *call,* or otherwise indicate a knowledge of).

TARGET DECK: The *deck* of cards the order of which the *subject* is *calling.*

TELEPATHY: *Extra-sensory perception* of the mental activities of another person. It does not include the *clairvoyant* perception of objective events.

TOUCH MATCHING: The technique in which the *subject* indicates his *call* by tapping or touching one of the five *key cards* while the *experimenter* places the *target card* thus called in front of the *key card* indicated.

TRIAL: A single *call* of a presented *symbol.*

X VALUE = CRITICAL RATIO: The observed *deviation* divided by the *standard deviation* or by the *probable error.*